SOAR-2

# State-of-the-Art Report

# PRACTICAL STATISTICAL ANALYSIS FOR THE RELIABILITY ENGINEER





Spring 1983



RAC

## Reliability Analysis Center

ROME AIR DEVELOPMENT CENTER



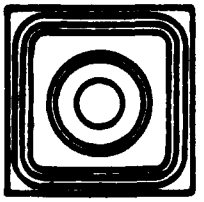
*THE RELIABILITY ANALYSIS CENTER IS A DOD INFORMATION ANALYSIS CENTER*

## ERRATA

Page 18. Figure 1e). The description under figure 1e) should read "r = -1".

Figure 3. Downtime is given on the abcissa as a proportion (not a % as stated).

Page 22. Second paragraph. "(on page 15)" should read "on page 19"



Reliability Analysis Center
A DoD Information Analysis Center

State-of-the Art Report

# PRACTICAL STATISTICAL ANALYSIS FOR THE RELIABILITY ENGINEER

Spring 1983

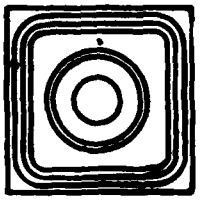
Prepared by:

Kieron A. Dey
IIT Research Institute

Under Contract to:

Rome Air Development Center
Griffiss AFB, NY 13441

Ordering No. SOAR-2



**The Reliability Analysis Center is a DoD Information Analysis Center, operated by IIT Research Institute under contract to the Rome Air Development Center, AFSC.**

The Reliability Analysis Center (RAC) is a Department of Defense Information Analysis Center sponsored by the Defense Logistics Agency, managed by the Rome Air Development Center (RADC), and operated at RADC by IIT Research Institute (IITRI). RAC is charged with the collection, analysis and dissemination of reliability information pertaining to parts used in electronic systems. The present scope includes integrated circuits, hybrids, discrete transistors and diodes, microwave devices, optoelectronics, and selected nonelectronic parts employed in military, space and commercial applications.

In addition, a System/Equipment Reliability Corporate Memory (RCM) is also operating under the auspices of the RAC and serves as the focal point for the collection and analysis of all reliability-related information and data on operating and planned military systems and equipment.

Data are collected on a continuous basis from a broad range of sources including testing laboratories, device and equipment manufacturers, government laboratories, and equipment users, both government and nongovernment. Automatic distribution lists, voluntary data submittal, and field failure reporting systems supplement an intensive data solicitation program.

Reliability data documents covering most of the device types mentioned above are available annually from RAC. Also, RAC provides reliability consulting and technical and bibliographic inquiry services which are fully discussed at the end of this document.

**REQUESTS FOR TECHNICAL ASSISTANCE AND INFORMATION ON AVAILABLE RAC SERVICES AND PUBLICATIONS MAY BE DIRECTED TO:**

**Harold A. Lauffenburger**
**IITRI/Reliability Analysis Center**
**RADC/RAC**
**Griffiss Air Force Base, NY 13441**
**Telephone: (315)330-4151**
**Autovon: 587-4151**

**ALL OTHER REQUESTS SHOULD BE DIRECTED TO:**

**Rome Air Development Center**
**RBE/Charles F. Bough**
**Griffiss Air Force Base, NY 13441**
**Telephone: (315)330-4920**
**Autovon: 587-4920**

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

## REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

| | |
|---|---|
| 1. REPORT NUMBER: SOAR-2 | 2. GOVT ACCESSION NO.: AD-A131270 |
| 3. RECIPIENT'S CATALOG NUMBER: | |
| 4. TITLE (and Subtitle): Practical Statistical Analysis for the Reliability Engineer | 5. TYPE OF REPORT & PERIOD COVERED: N/A |
| | 6. PERFORMING ORG. REPORT NUMBER: |
| 7. AUTHOR(s): K.A. Dey | 8. CONTRACT OR GRANT NUMBER(s): F30602-81-C-0299 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS: IIT Research Institute, Reliability Analysis Center, Griffiss AFB, NY 13441 | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: |
| 11. CONTROLLING OFFICE NAME AND ADDRESS: Rome Air Development Center, Griffiss AFB, NY 13441 | 12. REPORT DATE: March 1983 |
| | 13. NUMBER OF PAGES: 180 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): | 15. SECURITY CLASS. (of this report): |
| | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Available from the Reliability Analysis Center at $36 in the U.S. ($46 non-U.S.), prepaid. Also available from the National Technical Information Service (NTIS). No microfiche or microfiche-reproduced paper copies may be sold or otherwise distributed.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Statistics
Data Analysis
Parametric
Nonparametric
Reliability Data

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This text introduces a variety of standard statistical methods applicable to reliability data analysis. It is aimed at the non-specialist or manager and as such is primarily an applications guide. Methods addressed include analysis of variance, confidence interval estimation, goodness-of-fit tests, Weibull plotting, sampling inspection and regression and covariance analysis. Examples cover use of both parametric and nonparametric methods.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS


| | Page |
|---|---|
| ACKNOWLEDGEMENTS | xiii |
| INTRODUCTION | 1 |
| Objectives of This Text | 1 |
| Background | 1 |
| Two Cautions | 2 |
| A Corollary | 3 |
| Statistical Computing | 3 |
| Parametric and Nonparametric Statistics | 4 |
| CHAPTER 0 The Concept of Statistical Hypothesis Testing | 7 |
| Summary | 8 |
| Introduction | 8 |
| Inference | 10 |
| Lies, Damn Lies and Statistics | 11 |
| Significance | 12 |
| Conclusion | 13 |
| CHAPTER 1 THE CORRELATION COEFFICIENT | 15 |
| Summary | 16 |
| Covariance and Correlation | 16 |
| Significance of the Correlation Coefficient | 18 |
| Example | 18 |
| The Meaning of "Significant Correlation" | 20 |
| Model Assumptions | 21 |
| The Spearman Rank Correlation Coefficient | 22 |
| Linearity | 24 |
| Review | 25 |
| APPENDIX TO CHAPTER 1 | 26 |

TABLE OF CONTENTS (Cont'd)


| | Page |
|---|---|
| CHAPTER 2 ANALYSIS OF VARIANCE | 27 |
| Summary | 28 |
| What is Analysis of Variance? | 28 |
| Pardon? | 29 |
| How Does It Work? | 29 |
| Mathematical Model | 31 |
| Example 1 | 31 |
| How Long Does it Take? | 42 |
| Numerical Solution | 42 |
| What if the Model Assumptions are Violated? | 43 |
| The Kruskal-Wallis One-Way Method (Example 2) | 43 |
| Philosophy of Multiple Comparison Tests | 45 |
| Review | 45 |
| CHAPTER 3 CONFIDENCE INTERVALS | 47 |
| Summary | 48 |
| Introduction and Definition | 48 |
| At the Drop of a Hat - An Analogy | 49 |
| The Concept of a Confidence Interval for the Normal Distribution | 50 |
| Example | 53 |
| Confidence Intervals for the Mean of an Exponential | 56 |
| Another Example | 58 |
| Confidence Intervals for Parameters of the Weibull Distribution | 59 |
| Nonparametric Confidence Intervals on the Mean (For Complete Samples) | 60 |
| Concluding Remarks | 61 |
| CHAPTER 4 GOODNESS OF FIT TESTS | 63 |
| Summary | 64 |
| The Chi-Square $X^2$ Test | 64 |

TABLE OF CONTENTS (Cont'd)


| | Page |
|---|---|
| CHAPTER 4 GOODNESS OF FIT TESTS (Cont'd) | |
| Example I | 66 |
| The Kolmogorov-Smirnov Test | 69 |
| Example II | 70 |
| Comparison of the Two Tests | 71 |
| Review | 72 |
| CHAPTER 5 SAMPLING INSPECTION AND STATISTICAL Q.C. | 73 |
| Summary | 74 |
| Introduction | 74 |
| Evaluation of Batch Quality by Binomial Sampling | 76 |
| Producer and Consumer Risks | 79 |
| How Large Should the Sample Be? | 80 |
| LTPD Plans | 80 |
| Example | 82 |
| Small Batch Procedures | 82 |
| AQL Plans | 83 |
| Operating Characteristic Curves | 83 |
| Single, Double and Multiple Sampling Plans | 85 |
| Comparison of LTPD and AQL Schemes | 86 |
| Applications in the Electronic Component Industry | 86 |
| CHAPTER 6 WEIBULL PLOTTING AND ANALYSIS | 89 |
| Summary | 90 |
| Introduction | 90 |
| Weibull Plotting | 92 |
| Example | 96 |
| What if a Straight Line Does Not Result? | 99 |
| Optimization Methods | 100 |
| Physical Considerations | 102 |
| Further Methods | 102 |
| Review | 102 |

TABLE OF CONTENTS (Cont'd)


| | Page |
|---|---|
| CHAPTER 7 BAYESIAN STATISTICS | 103 |
| Summary | 104 |
| Introduction | 104 |
| The Controversy | 105 |
| The Likelihood Function | 107 |
| Numerical Example | 108 |
| Conclusions | 111 |
| CHAPTER 8 GRAPHICAL REGRESSION ANALYSIS | 113 |
| Introduction | 114 |
| Procedure | 114 |
| Restrictions | 116 |
| Uses | 116 |
| Transformations | 116 |
| CHAPTER 9 REGRESSION ANALYSIS AND LEAST SQUARES THEORY | 119 |
| Summary | 120 |
| Introduction | 120 |
| The Linear Model | 121 |
| Why the Linear Model? | 122 |
| Solution of the Linear Model - Least Squares Theory | 123 |
| How Good is the Fit? | 124 |
| How Good is Good? | 126 |
| Is There a Better Fit? | 128 |
| Numerical Example | 129 |
| Computer Solution | 133 |
| Scientific and Predictive Models | 133 |
| The Multiplicative Model | 135 |
| Example | 136 |
| Data Problems | 137 |
| Qualitative (Discrete) Models | 140 |
| Review | 140 |

TABLE OF CONTENTS (Cont'd)


| | Page |
|---|---|
| CHAPTER 10 A HOMOGENEITY TEST FOR NOISY, INCOMPLETE DATA | 143 |
| Summary | 144 |
| Data | 144 |
| Method | 145 |
| Data Analysis for Conventionally Censored Data | 146 |
| Artificially Grouped Data | 148 |
| Computer Outputs | 149 |
| APPENDIX TO CHAPTER 10 | 158 |
| Stat-10 Main Program | 159 |
| Subroutine RANNOS | 165 |
| Subroutine Histogram | 167 |
| Subroutine TPLOT | 171 |
| CHAPTER 11 COVARIANCE ANALYSIS | 177 |
| Summary | 178 |
| Introduction | 178 |
| Model | 178 |
| Transformations | 181 |
| Model Solution | 182 |
| Example | 182 |
| Review | 186 |
| CHAPTER 12 USE OF STATISTICAL TABLES | 189 |
| Summary | 190 |
| The Need for Tables | 190 |
| What is Tabulated? | 190 |
| CONCLUDING REMARKS | 195 |
| REFERENCES | 197 |
| ADDITIONAL RAC SERVICES | 201 |

LIST OF FIGURES


| | | Page |
|---|---|---|
| FIGURE 1: | CORRELATION | 17 |
| FIGURE 2: | SCATTER PLOT OF DOWNTIME AGAINST AGE | 20 |
| FIGURE 3: | DISTRIBUTIONS OF DOWNTIME AND AGE | 21 |
| FIGURE 4: | NON-LINEAR RELATION | 24 |
| FIGURE 5: | LINEAR RELATION | 25 |
| FIGURE 6: | HISTOGRAMS OF RAW DATA | 33 |
| FIGURE 7: | LOGARITHMIC TRANSFORMED HISTOGRAMS | 34 |
| FIGURE 8: | THE NORMAL DISTRIBUTION | 50 |
| FIGURE 9: | THE DISTRIBUTION OF THE MEAN OF A NORMAL | 51 |
| FIGURE 10: | DISTRIBUTION OF $LOG_{10}\ \lambda_o/\lambda_p$ | 54 |
| FIGURE 11: | PROBABILITY PLOT (NORMAL) FOR DATA | 55 |
| FIGURE 12: | THE EXPONENTIAL DISTRIBUTION | 57 |
| FIGURE 13: | GOODNESS OF FIT - HISTOGRAM AGAINST p.d.f. | 64 |
| FIGURE 14: | IDEAL O-C CURVE | 84 |
| FIGURE 15: | PRACTICAL O-C CURVE | 84 |
| FIGURE 16: | WEIBULL DENSITIES, HAZARDS AND CUMULATIVE DENSITIES | 91 |
| FIGURE 17: | WEIBULL PAPER | 94 |
| FIGURE 18: | WEIBULL PLOT | 98 |
| FIGURE 19: | RANDOM DIGITS PLOTTED ON WEIBULL PAPER | 101 |
| FIGURE 20: | PRIOR AND POSTERIOR DENSITIES | 111 |
| FIGURE 21: | GRAPHICAL APPROXIMATION TO A LEAST SQUARES SOLUTION | 115 |
| FIGURE 22: | EXAMPLES OF DATA TRANSFORMATIONS | 117 |
| FIGURE 23: | MINIMIZING THE SQUARED DEVIATIONS | 125 |
| FIGURE 24: | EFFECT OF GOOD AND BAD MODELS ON RESIDUAL VARIABILITY AND $R^2$ | 127 |
| FIGURE 25: | REGRESSION OF Y ON X | 130 |
| FIGURE 26: | RESIDUAL PLOT | 131 |
| FIGURE 27: | TYPES A, B AND C REGRESSIONS OF V ON R AND C IN THE V,C PLANE | 138 |
| FIGURE 28: | TYPES A, B AND C REGRESSIONS OF V ON R AND C IN THE V,R PLANE | 139 |

LIST OF FIGURES (Cont'd)


| | | Page |
|---|---|---|
| FIGURE 29: | OBSERVED MTTFs | 145 |
| FIGURE 30: | SIMULATED MTTFs | 146 |
| FIGURE 31: | BEHAVIOR OF $\hat{\theta}$ WITH r | 147 |
| FIGURE 32: | RESIDUAL ERROR | 185 |
| FIGURE 33: | MODEL FIT | 187 |
| FIGURE 34: | A GENERALIZED STATISTIC | 191 |
| FIGURE 35: | THE FAMILY OF DISTRIBUTIONS | 193 |

LIST OF TABLES


| | | Page |
|---|---|---|
| TABLE 1: | FACTORY DATA I | 19 |
| TABLE 2: | FACTORY DATA II | 22 |
| TABLE 3: | FAILURE VOLTAGES | 32 |
| TABLE 4: | ANALYSIS OF VARIANCE CALCULATIONS | 37 |
| TABLE 5: | ANOVA TABLE | 39 |
| TABLE 6: | AIRCRAFT DATA I | 43 |
| TABLE 7 | AIRCRAFT DATA II | 44 |
| TABLE 8: | NONPARAMETRIC CONFIDENCE LIMITS | 61 |
| TABLE 9: | HISTOGRAM GROUPING OF DATA | 67 |
| TABLE 10: | OBSERVED AND EXPECTED FREQUENCIES | 68 |
| TABLE 11: | OBSERVED AND EXPECTED CUMULATIVE FREQUENCIES | 70 |
| TABLE 12: | EXPECTED NUMBER OF DEFECTIVES PER SAMPLE | 77 |
| TABLE 13: | OBSERVED NUMBER OF DEFECTIVES | 78 |
| TABLE 14: | REPEATED SAMPLES | 78 |
| TABLE 15: | CONSEQUENCES AND RISKS | 79 |
| TABLE 16: | SAMPLING TABLE (Reproduced from MIL-M-38510E) | 81 |
| TABLE 17: | COMPARISON OF AQL AND LTPD | 86 |
| TABLE 18: | WEIBULL SHAPE PARAMETERS | 92 |
| TABLE 19: | RELATION BETWEEN $\beta$ and MTTF | 96 |
| TABLE 20: | DATA, RANKS AND PERCENTILES | 97 |
| TABLE 21: | REGRESSION DATA | 129 |
| TABLE 22: | REGRESSION STATISTICS | 132 |
| TABLE 23: | REFINED REGRESSION STATISTICS | 133 |
| TABLE 24: | COMPARISON OF THREE REGRESSION MODELS | 137 |
| TABLE 25: | FAILURE VOLTAGE DATA | 182 |
| TABLE 26: | LOGARITHMS OF FAILURE VOLTAGES AND CODED COVARIATE | 184 |

ACKNOWLEDGEMENTS

Little of this text presents original work; for the most part, it is simply a gathering together of useful techniques which are acknowledged by reference. Listed references cover the subject far more completely and elegantly than has been possible here.

Thanks to S.J. Flint (IITRI/Reliability Analysis Center) for inventive ideas and engineering support; J.L. Romeu (IIT Research Institute) for reviewing the statistical content; to Cliff Carroll, Helen Downing, Kim Flint and Sue Ziankoski who also contributed much to this text; and finally to the students of IIT Research Institute's "Semiconductor Training Course (1981/2)" on whose practical sessions the diode examples of Chapter 5 are based.

# INTRODUCTION

## Objectives of This Text

The increasing volume of statistical inquiries from users and from industry/government departments prompted the Reliability Analysis Center to publish this text. The nature of those inquiries indicated a need for a publication covering the following points:

- Introduce the non-specialist to statistical concepts with a minimum of mathematics
- Explain some popular methods applicable to practical reliability studies by example
- Acquaint the busy manager (and others) with some preliminaries so that the statistical content of technical reports may be understood
- Provide references for more advanced methods while at the same time indicating their potential

## Background

The origins of this text lie in a series of discussions and projects with colleagues of diverse disciplines. Common themes and problems emerged.

First, most engineers, computer scientists and technicians have had one or two courses in statistics but find the basic methods often unsuitable for real-world problems. The statistician can often solve real-world problems using advanced or esoteric statistics but such methods are not generally known to the non-specialist.

Second, many statistics texts (elementary and advanced) require the reader to have some prior interest in statistics and mathematics. Often such an interest is not enjoyed by those who need to analyze data, yet the need for formal statistical analysis is clear (or soon becomes so).

The most frequently recurring statistical questions and points of discussion at the Reliability Analysis Center have been the following (note that some are related):

- o What valid methods are available?
- o How do the methods relate to practical requirements?
- o Can I do them myself?
- o If I can't do them myself, how can I appreciate what the statistician might do for me with advanced methods?
- o Is there a quicker way?
- o Statistical methods always give me a numerical result, but how do I know it's the correct one? (In effect, was the right method properly applied?)
- o How can I understand statistical methods when used in technical reports?

To support advice given to colleagues in response to such questions, I began writing a series of "stand alone" statistical notes which could be referred to at later dates. These notes form the basis of individual chapters presented here. This text does assume a basic knowledge of elementary probability and statistics (e.g., probability density functions). Statistical tables are not reproduced here (they are extensively available in most statistics texts and as separate publications); however, their use is discussed in Chapter 12.

## Two Cautions

There are potential dangers in applying statistical methods from a user guide which are similar to those inherent in non-judicious use of

the many commercial software packages currently available. The following cautions are therefore stressed:

(i) This book is not pedagogic to the extent that a newcomer may learn statistics from it. It provides a summary, some examples, and some comments in the form of a user guide; however, a text this brief cannot convey the full implications of nor the vast volumes of work leading up to modern statistical methods. The reader wishing to learn statistics completely should refer to references given and possibly consider formal professional courses.

(ii) In practice, any given set of data is unlikely to be completely analyzed by a single method. Generally a number of different techniques will be applicable, each reviewing a different facet of the data, complementing each other, building up to a thorough analysis and sound conclusion. The stand-alone format of this text was chosen for convenience and ease of understanding, but it should be noted that applied statistical methods complement each other, and many of the methods described here are also strongly interrelated.

## A Corollary

Clearly, one answer to all this is that statistics is a profession where the services of a professional should be sought. However, it is also clear that many methods can be applied by the non-specialist and, in so doing, improve the statistician/engineer interaction, leaving the statistician free to tackle non-standard or original problems.

## Statistical Computing

Major improvements in statistics have been possible by the availability of computer solution. Fields such as time series analysis

have advanced into areas which would not have been feasible manually. Unfortunately, the myriad of commercial programs also available for standard (hitherto manual or semi-automated) methods have instilled some misapplication in science and industry, for example by applying a standard method to a nonstandard data structure. Data has to be understood to be properly analyzed, and a simple way to understand data is to analyze it manually. Though strategic access to a computer is also essential (at the correct stages of an analysis), the time "lost" by manual analysis (at least initially) is far less than one might believe.

A library of subroutines (to be programmed for each individual analysis) can be far more powerful than the standard statistic packages which ask only a set list of questions about the data and provide answers meaningfully or not.

## Parametric and Nonparametric Statistics

Most of the well-known statistical methods are based on underlying distributions which the data are found to follow (e.g., normal, exponential, Poisson, etc.). Statistical models and tests are then constructed using distribution theory, and the data is evaluated using the final model or test. These are called parametric methods. The use of these methods does depend on the distribution being known, that the measure of the data (e.g., hours until failure) is reasonably precise, and that a good-sized sample is available. If any or all of these conditions do not hold, then, strictly, parametric methods are invalid and it may be preferable to use nonparametric methods.

Nonparametric methods evaluate the probability of sequences and combinations of data independent of distributional assumptions (except that the data be reasonably symmetrically dispersed). They are generally weaker tests than their parametric counterparts. They are

invaluable where the measure used is imprecise (for example, if there is noise on the measure or the measure is merely an ordering of failure rates). It should also be noted that, far from being a last-ditch alternative, nonparametric methods offer an attractive approach to data analysis. Comparative examples of both styles of data analysis are given in this text.

# CHAPTER 0

## The Concept of Statistical Hypothesis Testing

CHAPTER 0

# The Concept of Statistical Hypothesis Testing

## Summary

The concepts introduced in this chapter should be understood before using any statistical methods. It describes the philosophy commonly referred to as the Neyman-Pearson theory. Lack of understanding of this theory leads to many misinterpretations and invalid conclusions. It is as important to those reading technical reports as it is to those writing them.

## Introduction

The purpose of any statistical text is to assess without bias some aspect of a set of data. In the 1920s and 1930s, Jerzy Neyman and Egon Pearson defined a theory which was to form the basis of modern statistical method. Very roughly stated, the theory says that in order to make unbiased inferences from data it is necessary to decide on some basic premise which the experimenter (or analyst) believes most likely to reflect the status quo. The basic premise is made before analyzing the data and is based on an understanding of the physical principles involved in the experiment or situation from whence the data came. This basic premise is termed the null hypothesis. The data is then analyzed and the null hypothesis reviewed in the light of the data. This is done by evaluating the probability of observing the data given that the null hypothesis is true. If the evaluated probability is significantly small then we reject the null hypothesis. If we reject the null hypothesis then we conclude in favor of some alternative hypothesis which must also be defined in advance of analyzing the data. Statistical convention uses the terminology $H_0$ for the null hypothesis and $H_1$ for the alternative.

This type of data analysis is termed "hypothesis testing." The philosophy of hypothesis testing gives statistical methods their objectivity. A useful analogy is the judicial system, where the null hypothesis is that the defendant is innocent against the alternative that he is guilty and the hypothesis of innocence is reviewed in the light of data (evidence). A jury is used to assess probability of innocence given the evidence. The judicial system is thus designed to be unbiased, and statistically it is. Note that two risks of error (to the defendant and to the court) exist for individual cases. Statistical convention defines this as Types I and II error.

A table of consequences for the risks of error is included in Chapter 5 using an example of inspection sampling for a semiconductor production process.

The risk of Type I error is conventionally denoted by $\alpha$, while the corresponding risk of Type II error is denoted by $\beta$. It has become accepted practice in industrial work to construct statistical tests by fixing $\alpha$ at some required level. However, the $\beta$ risk is also important, and $(1 - \beta)$ defines the power of the test: the power to reject data which in fact resulted from the alternative hypothesis being true. Thus, the power of a test is (crudely) its ability to reject the null hypothesis when in fact the alternative is true.

Statistical tests which evaluate whether a null hypothesis is true (against the alternative hypothesis that the null is untrue) are termed two-sided tests. One-sided tests evaluate a null hypothesis against the alternative hypothesis that the parameter being tested exceeds the null hypothesis. Thus, two-sided tests evaluate equality, while one-sided tests evaluate inequality.

For example, a two-sided test on the correlation between two variables might be defined on the null hypothesis that the correlation coefficient $r = 0$, against the alternative hypothesis that $r \neq 0$. The

null hypothesis is then evaluated using some test on the correlation coefficient, such as the one described in Chapter 1. In contrast, a one-sided test on r might be set up with the null hypothesis that $r = 0$ (as before) against the alternative hypothesis that $r < 0$. In fact this is shown by example in Chapter 1. This example uses the correlation coefficient r. In general, there are a variety of test statistics used to evaluate various aspects of given sets of data, examples follow in later chapters.

One-sided tests have greater power than two-sided tests but it is important to be sure that the inequality not being tested is either impossible or of no consequence.

## Inference

Inference is the process of drawing conclusions from data. Depending on the particular statistical method employed (e.g., analysis of variance, t-test, etc.) various assumptions are made about the structure of the data (e.g., normality, independence, etc.). If all these assumptions hold, then we can be confident (within the defined risk specified for the particular test) that any statistical inferences are correct. Any departure from the assumptions needed for the particular statistical method employed will increase the risk of an incorrect inference. In the judicial system, a suitable analogy would be as follows: the assumption is made that witnesses do not confer among themselves; if they did and if a trial (statistical test) proceeded regardless, there would be an increased risk of incorrect conclusion due to the "mob effect" in the witnesses' testimony. Returning to statistics, the normality or independence assumptions may not hold.

Thus, unless all assumptions are rigorously checked, there is potentially an increased risk of making incorrect inferences. Thus, when a "statistically significant" conclusion is reached, it may or may

not be the correct scientific conclusion. Its validity can only be checked by rigorous inspection. Statistical inference is not a straightforward issue nor is it an applied mathematical method (although mathematics is used in many of the methods). Adherence to the Neyman-Pearson theory as outlined earlier will considerably reduce the risk of error and allow what risk there is to be assessed.

## Lies, Damn Lies and Statistics

Part of the reason for this catchy phrase is due to the problems outlined above. But why is it that we don't hear this phrase in using radar and other complex systems which use statistical methods in making decisions about, for example, the existence or otherwise of a target? Probably because of the measurement system. In a radar, nobody doubts the validity of a set of data (signal) as measured by an aerial (it is interesting to note that sampling rates are carefully assessed to prevent wrong conclusions, e.g., aliasing). Thus, any inferences (detections) based on such data are considered quite valid (and rightly so), and the statistics used are identical to those used in data analysis. However, in applying statistics to measuring, for example, economic performance or reliability, things are not quite as neatly defined. Does number of failures per year measure reliability? Possibly not, if the equipment is a car and the user is a stuntman (bear in mind we won't know how well he drives the car and will take his failure rate as indicative of inherent reliability if we are not very careful). Politicians tend to oversimplify the economist's work, so it is not surprising that a statistical analysis here will be controversial. Clearly there are plenty of bad statistics in print, but this is not all of the problem. For an academic appraisal of the problems of measures see Reference 2. The final problem giving rise to invalid conclusions is lack of randomization in sampling. This is very important and very difficult.

Turning the issue around, is it possible that a correct conclusion can be reached and yet still receive the same reaction as defined in the

heading of this section? The answer is most certainly "Yes"; the human brain is often very biased, sometimes irrationally so. For example, development of a phobia is intellectually unreasonable and represents a biased assessment of risk. In the case of fear of flying, the data is often hundreds of successful safe flights and yet the victim focuses on subjective, imaginary data attaching a greater probability to crashing than to landing safely. Clearly this is an extreme case but it does illustrate the problem and there are degrees of bias of lesser severity which are indetectable or vary from person to person. This is the most difficult part of the statistician's work.

## Significance

A couple of pages ago, we discussed how to set up a statistical test to evaluate some hypothesis without bias. It is also necessary to have some objective means of deciding when to reject the null hypothesis, and when not to. The decision criteria is defined by the significance level of the test. Significance is a measure of how extreme (compared to the "norm") our data is. This measure is simply the probability of observing a result as extreme (or more so) as the one observed. The more extreme the result the lower the measured probability; hence, we typically use the following table to define whether a result is "significant" or not. When we say "significant" we really mean "significantly different from the norm."

| Significance Level | Description |
|---|---|
| 10% | Weakly significant |
| 5% | Significant |
| 1% | Highly significant |
| 0.1% | Very highly significant |

A less significant result than 10% (e.g., 30%) would normally be taken as insignificant since we have a high probability of it happening and hence it is not a surprising result. i.e., it doesn't significantly

differ from the norm. Usually the $\alpha$ = 5% level of significance is used for scientific work unless there are any special conditions. It is necessary to quote the significance level $\alpha$ in advance of analyzing a set of data.

In practice, it is tedious to compute the probability every time a test is carried out, and compare it to $\alpha$. Instead, values of the test statistic being used are tabulated (e.g., Ref. 3). The value of that statistic, corresponding to a significance $\alpha$, is termed the critical value. More discussion on tables is given in Chapter 12.

Hence, the process of setting up and using a hypothesis test may be summarized as follows:

(a) Define $H_0$ (the null hypothesis)
(b) Define $H_1$ (the alternative hypothesis)
(c) Define the test statistic which will be used to evaluate the data against the null hypothesis
(d) Define $\alpha$
(e) Look up the critical value of the test statistic at the $\alpha$ significance level
(f) Evaluate the test statistic for the data in hand
(g) Compare the test statistic value to the critical value. Reject $H_0$ if the test statistic (from the data) exceeds the critical value

## Conclusion

The concept of hypothesis testing is applicable in using most of the methods described in this text. Asking the right questions about a given data set (i.e., defining $H_0$) is as important as applying the methods correctly. A note as short as this is inadequate to fully define statistical hypothesis testing but it does provide a review of the basics. References for deeper discussions are included in the bibliography (References 1, 4, 5).

# CHAPTER 1

## The Correlation Coefficient

CHAPTER 1

# The Correlation Coefficient

## Summary

Although the correlation coefficient is commonly used, it is not always understood that its significance may be calculated and has been tabulated. This chapter covers derivation of r and its sampling distribution. Significant correlations tend to be numerically less than one might expect. Use of the correlation coefficient (and its significance tables) may not always be valid. This chapter also discusses such cases and recommends alternatives.

## Covariance and Correlation

The covariance of a pair of random variables X and Y is a measure of the extent to which they are related to one another. It may be estimated from a sample of data by:

$$\text{Cov. } (X, Y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})\,(y_i - \bar{y})}{n} \qquad (1)$$

where

$x_i$ is the $i^{th}$ observation in X
$y_i$ is the $i^{th}$ observation in Y
n is the total number of observations

Thus, the covariance may be evaluated from a set of paired observations $(x_1, y_1)$, $(x_2, y_2)$ ... $(x_n, y_n)$ in X and Y.

The regularly used correlation coefficient r is a standardized covariance defined as:

$$r = \frac{\text{Cov}\ (X,Y)}{S_x, S_y} \qquad (2)$$

where

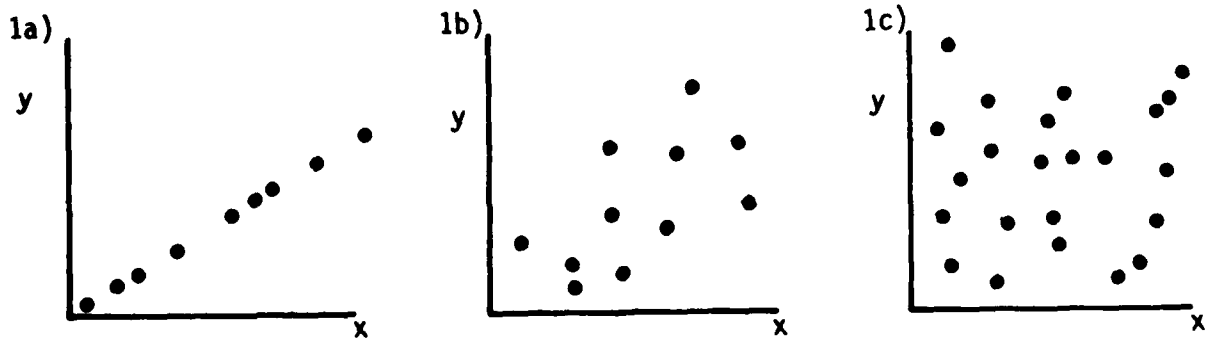
$s_x$ is the standard deviation of X estimated from the sample $(x_1 \ldots x_n)$

$s_y$ is the standard deviation of Y as estimated from the sample $(y_1 \ldots y_n)$

So

$$s_x = \left[ \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{(n-1)} \right]^{\frac{1}{2}} \qquad \text{and } s_y = \left[ \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{(n-1)} \right]^{\frac{1}{2}}$$

r will be found to vary from -1 to +1 according to whether the data is perfectly correlated (as in Figure 1a)



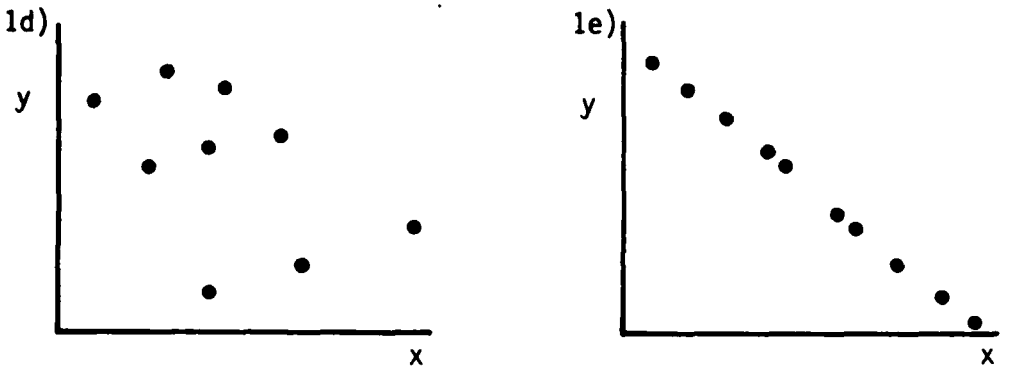

PERFECT (LINEAR) CORRELATION
r = 1

SMALL CORRELATION
0 < r < 1

NO CORRELATION
r = 0

FIGURE 1: CORRELATION

through to perfectly negatively correlated (Figures 1b to 1e).




SMALL NEGATIVE CORRELATION
$-1 < r < 0$

PERFECT NEGATIVE
LINEAR CORRELATION
$r = 1$

FIGURE 1: CORRELATION (Cont'd)

In practice, data is usually not perfectly correlated due to random fluctuation and measurement error. The purpose of the correlation coefficient is to measure how far from the two extremes (of -1 and 1) our data actually is and to apply a significance test to statistically evaluate whether any correlation exists.

## Significance of the Correlation Coefficient

It is not intuitively obvious how large an r value has to be to indicate a significant correlation and so the sampling distribution of r is required. Exact derivation is difficult but approximately $r\sqrt{n-2} \,/\, \sqrt{1-r^2}$ has a t distribution (where n is the number of data pairs) if X and Y are distributed as a bivariate normal. Significance values of r are tabulated in standard tables, as illustrated in the following examples.

## Example

The following data on a number of similar factories was obtained. The data concerned downtime and included recordings of a number of other variables. Management needed to know whether the age of the factory and the downtime were correlated.

The data was as follows:

TABLE 1: FACTORY DATA I

| Factory Index # | Age (years) | Downtime (% per year) |
|---|---|---|
| 1 | 7.88 | 28.0 |
| 2 | 7.01 | 30.7 |
| 3 | 4.97 | 9.7 |
| 4 | 4.74 | 18.1 |
| 5 | 6.34 | 18.2 |
| 6 | 4.59 | 28.1 |
| 7 | 11.39 | 12.2 |
| 8 | 10.11 | 14.1 |
| 9 | 8.18 | 9.6 |
| 10 | 8.32 | 16.7 |
| 11 | 7.74 | 16.1 |
| 12 | 7.00 | 15.8 |
| 13 | 9.39 | 7.1 |
| 14 | 9.28 | 8.5 |
| 15 | 10.93 | 14.2 |
| 16 | 1.11 | 30.9 |
| 17 | 8.18 | 13.5 |
| 18 | 7.68 | 15.7 |

First, we set up the null hypothesis that there is no relation between age and downtime against the alternative that there is a negative correlation (i.e., this is a one-tailed test).

On a scatter diagram, the data looks like Figure 2 on the following page. Some negative correlation is indicated but it is hard to be sure.

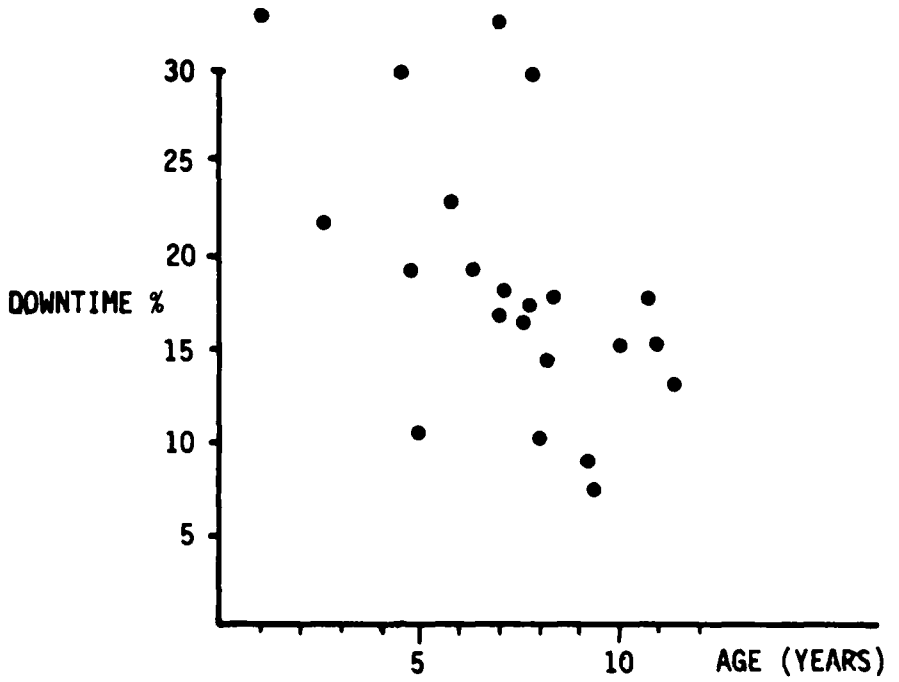



FIGURE 2: SCATTER PLOT OF DOWNTIME AGAINST AGE

From (1) the correlation coefficient is evaluated as the covariance of age and downtime over the product of their standard deviations. Calculations yield:

Cov (age, outage) = -10.38

The age and downtime standard deviations are respectively estimated as 2.5139 and 7.5208. Consequently, r is found to be about -0.549. (The full calculations are shown in the appendix to this note; these computations are generally carried out on a computer; little is gained by hand-calculation.)

Now, from statistical tables (Ref. 3) the critical value of the correlation coefficient is found to be 0.4 at the 5% level and the null hypothesis is rejected since our value exceeds the critical value.

The Meaning of "Significant Correlation"

The fact that age is found to be significantly correlated with downtime does not necessarily mean that the age reduces the downtime.

It may well be that the older factories are smaller and for some reason related to size fail less often. The question as to whether downtime improves with age (because of maturity) can only be answered by an understanding of the factories' operation and possibly an analysis of data on other variables as well. This would probably require either analysis of variance, multiple regression, or both, as described in Chapters 2, 9 and 11. Thus, correlation does not necessarily imply dependency: where there is no direct dependency the correlation is termed spurious.

## Model Assumptions

Recall that the calculation of the significance assumes the two variables to be bivariate normally distributed. Projection of the age/downtime data indicates that such an assumption is reasonable though far from perfect (see below) but this will not always be the case. In addition, variables are not always precisely measurable.

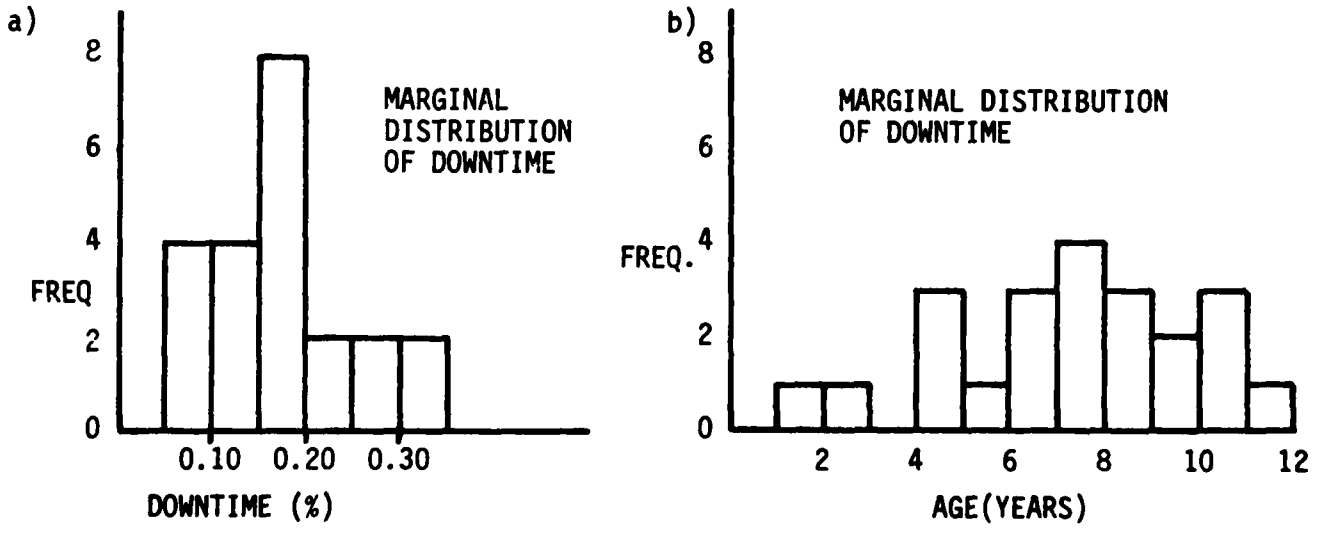



FIGURE 3: DISTRIBUTIONS OF DOWNTIME AND AGE

It is therefore desirable to have some alternative means of measuring correlation and in fact several different nonparametric methods are available. One of the simplest to apply is the Spearman rank correlation coefficient which is described below. However, there are other methods which may be more powerful under given conditions.

## The Spearman Rank Correlation Coefficient

Reconsider the data in Table 1 (on page 15). The factories may be ranked by age or by downtime (with respect to their index numbers) as in Table 2 below:

TABLE 2: FACTORY DATA II

| Rank | Age | Downtime |
|---|---|---|
| 1 | 16 | 13 |
| 2 | 6 | 14 |
| 3 | 4 | 9 |
| 4 | 3 | 3 |
| 5 | 5 | 7 |
| 6 | 12 | 17 |
| 7 | 2 | 8 |
| 8 | 18 | 15 |
| 9 | 11 | 18 |
| 10 | 1 | 12 |
| 11 | 9, 17 | 11 |
| 12 | | 10 |
| 13 | 10 | 4 |
| 14 | 14 | 5 |
| 15 | 13 | 1 |
| 16 | 8 | 6 |
| 17 | 15 | 2 |
| 18 | 7 | 16 |

So, for example, factory 16 is the youngest, has the highest downtime, and therefore, is ranked first (with respect to age) and last (with respect to downtime). Note that where ties exist (e.g., factories 9 and 17 are both 8.18 years old) the mean of the next two ranks is used (e.g., rank for 9 and 17 by age = (11 + 12)/2 = 11½). There is theoretical justification for doing this.

Now if the two rankings (by age and by downtime) were identical then we would conclude that there is a perfect (by this measure) correlation between age and downtime. If the two rankings were identical but reversed, the correlation would be perfectly negative. However, it is not easy to see how far this set of data is from either of the two extreme conditions (or from the obvious alternative that the correlation is zero).

Spearman's rank correlation coefficient measures correlation between two such rankings and is given by:

$$\rho = 1 - \frac{6 \sum_{j=1}^{m} d_j^2}{m^3 - m} \qquad (3)$$

where

m is the number of pairs of observations

$d_j$ is the deviation between the two ranks for a given observation

So, for example, $d_1$ is the difference between the two rankings of factory #1, i.e.,

$$d_1 = 10 - 15 = -5$$

Computing $d_j$ for all 18 factories gives

$$\sum_{j=1}^{m} d_j^2 = \big((-5)^2 + (-10)^2 + (0)^2 + (-10)^2 + (-9)^2 + (-14)^2 + 13^2 + 9^2 + 8.5^2 + 1^2 + (-2)^2 + (-4)^2 + 14^2 + 12^2 + 9^2 + (-17)^2 + 5.5^2 + (-1)^2\big) = 1586.5$$

So applying formula (2) gives:

$$\rho = 1 - \frac{6(1586.5)}{18^3-18} = -0.637$$

A negative correlation is indicated but it is still necessary to determine whether or not it is significant. In fact, the critical value of Spearman's rank correlation coefficient is found from tables (Ref. 7) to be 0.399 at the 5% level. Under the same null hypothesis as before, we reject that there is no correlation. Negative correlation between age and downtime is therefore indicated.

The agreement between the parametric and nonparametric methods is good. The latter uses less information but has consequently less power.

Linearity

Suppose a set of data in X and Y looks like this:


FIGURE 4: NON-LINEAR RELATION

Clearly there is a strong correlation though it is not linear. Application of the parametric correlation coefficient given in (1) would be invalid since it may only be used to detect linear correlation. For such cases there are two options:

(i) Apply the nonparametric method, e.g., Spearman's $\rho$ as given in (2), since it is unrestricted by linearity, requiring only an ordering

(ii) Apply a transformation to X or Y so that the correlation is linear in transformed scale as below

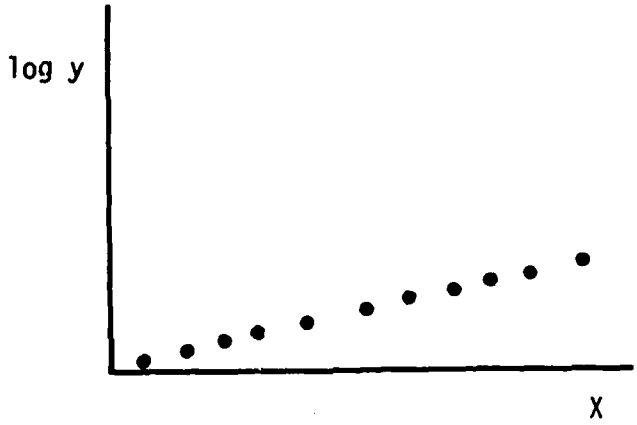



FIGURE 5: LINEAR RELATION

Review

The relation between two variables may be assessed and tested for significance using a correlation coefficient. The parametric method (r) assumes that the variables have a bivariate normal distribution, that the relation is a linear one, and that the variables are measured reasonably well. If any of these assumptions are violated then a nonparametric method is preferred; Spearman's rank correlation coefficient is shown by example. Nonlinear correlations may be assessed with r by using a transformation. A significant correlation does not necessarily imply that one variable is dependent on the other. It may be that the relation is a result of a third variable to which both are related though for different reasons.

APPENDIX TO CHAPTER 1

## Correlation of r

| i | $x_i$ | $y_i$ | $(x_i - \bar{x})$ | $(y_i - \bar{y})$ | $(x_i - \bar{x})(y_i - \bar{y})$ |
|---|---|---|---|---|---|
| 1 | 7.88 | 28.0 | 0.3889 | 10.9333 | 4.2520 |
| 2 | 7.01 | 30.7 | -0.4811 | 13.6333 | - 6.5590 |
| 3 | 4.97 | 9.7 | -2.5211 | -7.3667 | 18.5722 |
| 4 | 4.74 | 18.1 | -2.7511 | 1.0333 | - 2.8427 |
| 5 | 6.34 | 18.2 | -1.1511 | 1.1333 | - 1.3045 |
| 6 | 4.59 | 28.1 | -2.9011 | 11.0333 | -32.0087 |
| 7 | 11.39 | 12.2 | 3.8989 | -4.8667 | -18.9748 |
| 8 | 10.11 | 14.1 | 2.6189 | -2.9667 | - 7.7695 |
| 9 | 8.18 | 9.6 | 0.6889 | -7.4667 | - 5.1438 |
| 10 | 8.32 | 16.7 | 0.8289 | -0.3667 | - 0.3040 |
| 11 | 7.74 | 16.1 | 0.2489 | -0.9667 | - 0.2406 |
| 12 | 7.00 | 15.8 | -0.4911 | -1.2667 | 0.6221 |
| 13 | 9.39 | 7.1 | 1.8989 | -9.9667 | -18.9258 |
| 14 | 9.28 | 8.5 | 1.7889 | -8.5667 | -15.3250 |
| 15 | 10.93 | 14.2 | 3.4389 | -2.8667 | - 9.8583 |
| 16 | 1.11 | 30.9 | -6.3811 | 13.8333 | -88.2717 |
| 17 | 8.18 | 13.5 | 0.6889 | -3.5667 | - 2.4571 |
| 18 | 7.68 | 15.7 | 0.1889 | -1.3667 | - 0.2582 |
| | | | | $\Sigma(x_i - \bar{x})(y_i - \bar{y})$ = | -186.7974 |

$\bar{x}$ = 7.4911

$\bar{y}$ = 17.0667

n = 18

$$S_x = \left[\frac{\Sigma(x_i - \bar{x})^2}{n-1}\right]^{\frac{1}{2}} = 2.5139 \qquad \text{Cov } (X,Y) = 10.3776$$

$$S_y = \left[\frac{\Sigma(y_i - \bar{y})^2}{n-1}\right] = 7.5208 \qquad r = \frac{\text{Cov}(X,Y)}{s_x \cdot s_y} = -0.549$$

# CHAPTER 2

## Analysis of Variance

CHAPTER 2

# Analysis Of Variance

## Summary

Analysis of variance (ANOVA) is an extremely powerful technique which is given extensive treatment in the literature, little of which is comprehensible to the layman, unfortunately. This note outlines how to do such an analysis in very simple terms. The interested reader is strongly recommended to check the theory (References 4, 5) if he hopes to fully understand the method and its underlying assumptions. Basically, the method separates and evaluates the effect of each of a number of variables (e.g., temperature, complexity) which simultaneously influence some dependent variable (such as failure rate).

Since the standard assumptions are often not valid for field reliability data both the classical and nonparametric alternatives are given. Having read this note, the reader may still find it easier to farm out his work to a specialist since the method is one of the more complicated in modern statistics, but at least he should be able to follow the broad principles and understand such analyses when used in technical reports.

## What is Analysis of Variance?

Analysis of variance is the separation of the variance attributable to one group of causes from the variance attributable to other groups of causes. The significance of differences between such separated groups is determinable using the partitioning theorem of the $\chi^2$ distribution and hence applying an F test of significance. In this way the effect of one or more causes (or factors or treatments) may be evaluated. Consequently it may be possible to identify factors which degrade reliability by carrying out analysis of variance with failure rate (or similar) as the dependent variable.

## Pardon?

Consider the analogy of a fly buzzing about in a car. Suppose we observe this fly at a speed of 5 mph relative to the car which is doing 30 mph in the same direction, then the fly's speed is 35 mph relative to the earth. Further suppose that the earth orbits at 65,000 mph and, all the more remarkable, travels in the same direction as the car and the fly. Now the fly's speed is 65,035 mph relative to the sun.

Now suppose we were told only that this fly was doing 65,035 mph relative to the sun; we would go away and analyze the relative velocities and hence the "cause" or "causes" of the total.

Analysis of variance is similar to being told the fly's speed relative to the sun and making an objective decision as to whether the fly, car and earth speeds are the same or not. In the example, this is easy to tell, but if we had ten variables in three dimensions all operating in different directions, it would be very hard to sort out what was happening.

In analysis of variance, we look at variances in the data to draw inferences about the original velocities. Instead of velocities, though, we consider whatever factor(s) is most useful.

## How Does It Work?

If we have a set of failure rate data together with data on each of a number of other discrete variables (which may have contributed to the failure either separately or in combination) then it may be required to determine which of the variables "caused" the failures. For the reasons indicated above, the answer will not be immediately obvious, and any attempt to do this intuitively could be seriously misleading.

Thus, analysis of variance is an algorithm for mathematically separating the effects (on reliability) of each of a number of (or combinations of) discrete variables. An example might be failure rate data in each of four environments. Thus a part of the analysis would include a determination of whether or not the discrete variable of environment affected reliability.

The algorithm works essentially by comparing the variability between the four environments to the overall noise in the data (which is termed "residual variation"). If the variability between environments is greater than the noise, then the indication is that environment does have an effect on reliability. The between-environments variability is objectively compared to the residual variability by using a statistical test (the F test). The algorithm simultaneously evaluates the effects of as many other variables as are required given certain constraints. Variability is assessed using sums of squared deviations (SSD) of subsets of the data about the group means (to measure the noise) and the SSD of sets of group means about the overall mean (to measure the effect of each variable). Interactions are assessed similarly, though with a little more computational complexity.

Following the guidelines set out in Chapter 0, we generally set up the null hypothesis that all the variables have no effect and then try to disprove it, wholly or partially.

Interactions between variables are also analyzed. An example of interaction might be humidity and vibration, with respect to reliability. The combined effect of the two might be greater than the sum of the individual effects; hence there is an interaction. In practice, combinations of three, four, and umpteen factors may exist. Such higher-order interactions are possible although harder to diagnose.

## Mathematical Model

The underlying mathematical model and solution are well-defined in Reference 27. Essentially, the model assumes that any given failure rate observed under a given set of conditions (of a given set of discrete variables) is an additive function of a numerical realization of these conditions. Thus, considering again the environment example, each of the four environments may add respectively 0, 1.1, 2.3 and 3.9 failures per $10^6$ hours to some base failure rate (expected under neutral conditions of zero stress). Extension to two, three and subsequent variables follows.

But what if the effect is not additive (and intuitively, one would expect reliability not to be)? Experience in this type of work suggests that a multiplicative effect (i.e., each different environment adds a constant proportion rather than a constant amount) is appropriate. Multiplicative models are conveniently accommodated simply by taking logarithms. This is illustrated in the example later. Other transformations may be necessary for given data sets.

In addition, the model assumes that the data within each set of conditions is normally distributed with the same variance. This assumption applies after transformation, if there is one. In practice, the method is quite robust to minor deviations from the normality assumption.

It is also assumed that the sample is unbiased with respect to the variables being analyzed. Parametric procedures exist for biased, unbalanced or incomplete samples but their use is outside the scope of this text. Instead, a nonparametric alternative is suggested later.

## Example 1

RAC recently carried out a series of microcircuit tests to

determine their susceptibility to electrostatic discharge (ESD). The full program of testing was reported in Reference 6. One of the tests was to determine the applied electrostatic voltage at which parts failed. The test was carried out independently by two separate test houses (on two different samples of components). Also two different microcircuit types (a hex inverter and a quad D flip flop) were to be evaluated under two different test circuit configurations. One configuration was for all pins tied together and grounded (APTTG) and the other left the pins floating (APTTF). Consequently the objective of the test results analysis was to answer the following questions:

- o Do the two microcircuits exhibit the same failure voltage?
- o Did the different testers obtain the same results within the bounds of experimental error?
- o Were the failure voltages affected by test circuit configuration?
- o Was there any evidence that the testers used different conditions?

The whole problem is readily handled by analysis of variance. The results of the test are summarized in Table 3.

TABLE 3: FAILURE VOLTAGES

| | Hex Inverter | | Quad D Flip Flop | |
|---|---|---|---|---|
| | APTTG | APTTF | APTTG | APTTF |
| Tester I | 200 | 1800 | 1800 | 2200 |
| | 200 | 800 | 700 | 1400 |
| | 200 | 700 | 500 | 2200 |
| | 400 | 1800 | 1800 | 2200 |
| | 600 | 1400 | 2200 | 3000 |
| | 500 | 1400 | 2600 | 2200 |
| Tester II | 400 | 3000 | 500 | 4600 |
| | 400 | 2200 | 700 | 4600 |
| | 400 | 2200 | 3400 | 4600 |
| | 300 | 1400 | 300 | 2200 |
| | 400 | 900 | 400 | 3000 |
| | 400 | 900 | 800 | 4600 |

This data is sketched in Figure 6 as a series of histograms.

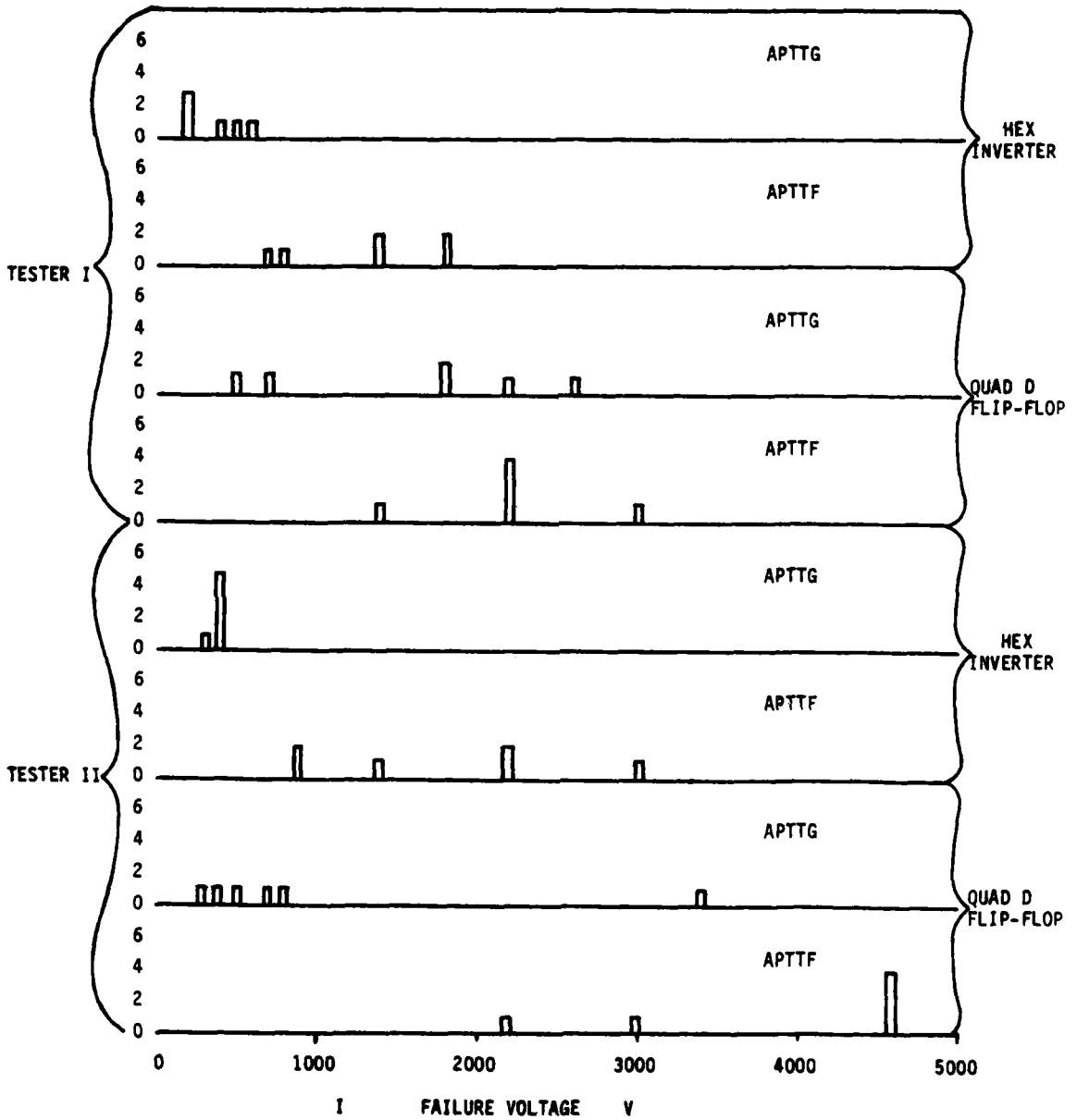



FIGURE 6: HISTOGRAMS OF RAW DATA

Figure 6 clearly shows that the variance is not the same for each histogram and that variance actually increases with failure voltage. This is an important clue that a logarithmic transformation is necessary to stabilize the variance. Figure 7 shows the same data on a logarithmic scale. Strictly, the transformed data should be checked for normality in each histogram (probably with a Kolmogorov-Smirnov goodness-of-fit test, as described in Chapter 4) and equality of variance (probably with Bartlett's test, Reference 5) but such detail is omitted here. Essentially we can see that the data has now reasonably equal variance and is normally dispersed. (Statisticians call this homoscedasticity.)

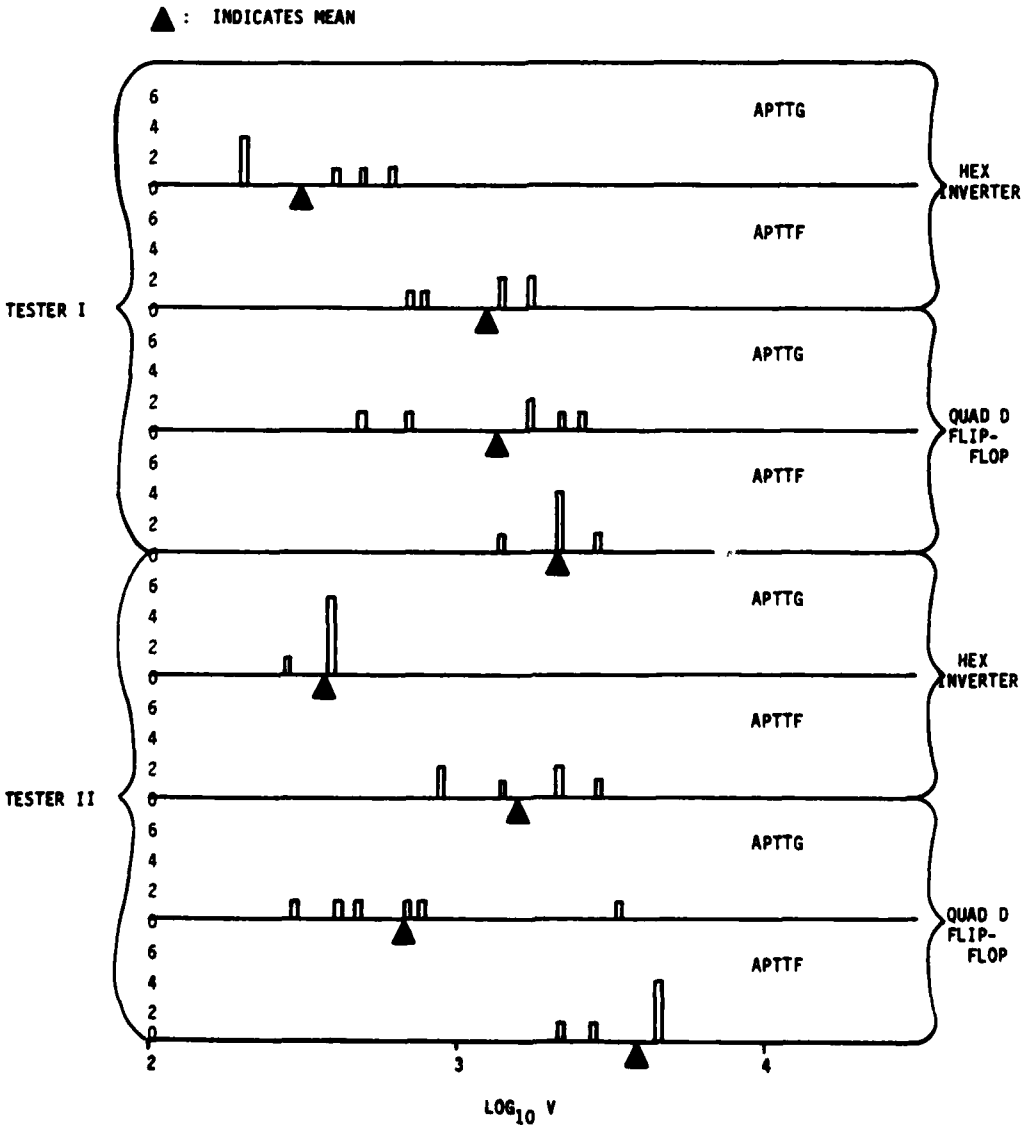



FIGURE 7: LOGARITHMIC TRANSFORMED HISTOGRAMS

Considering Figure 7, we can see that APTTF appears to induce higher failure voltages than APTTG and that the quad D flip flop appears to exhibit higher failure voltages than the hex inverter. Further conclusions are not apparent. Analysis of variance removes the words "appears to" and replaces them with an objective statistical decision, simultaneously extending to further conclusions which are difficult or misleading visually.

An analysis of variance procedure is now applied to the transformed data as follows (the purpose of each step is labelled in italics):

(i) *Purpose: Define $H_0$ and $H_1$ and significance ($\alpha$).* Set up the null hypothesis that none of the variables (either individually or in combination) affect failure voltage against the alternative that at least one variable affects failure voltage. The significance level is set at the standard 5%.

(ii) *Purpose: Calculate the total sum of squares.* Calculate the sum of the squared deviations from the overall mean. This can be done by the arduous method of first calculating the mean, subtracting it from each observation in turn, squaring each result, and summing. A quicker way, using a calculator which has statistical functions, is to calculate the variance and multiply it by n-1 (where n = the number of data points). Then since

$$\text{Variance estimate} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n - 1}$$

where $x_i$ is the $i^{th}$ data point
and $\bar{x}$ is the overall mean of the $x_i$

Then (n-1) (variance estimate) = $\Sigma(x_i - \bar{x})^2$ which is what we need. This quantity is called the <u>total sum of squares</u>.

For the data in hand, the overall mean is found to be 3.0327 and the total sum of squares is 7.6643.

The total sum of squares is used to find other sums of squares by subtraction.

(iii) *Purpose: Calculate the SSD for each group of data.* The sum of squared deviations for any group of data is given by subtracting the overall mean from the mean of that group, squaring the result and multiplying by the number in the group. Thus, if we want the SSD for the Tester I data, it is found by:

$$SSD_{\text{Tester I}} = (\text{Group mean} - \text{Overall mean})^2 \times 24$$

So

$$SSD_{\text{Tester I}} = (3.0141 - 3.0327)^2 \times 24 = 0.00833$$

and the procedure is repeated for all groups. Results are given in Table 4.

(iv) *Purpose: Calculate the interaction SSD for each pair of variables.* Interactions between variables now have to be investigated. For this example, it is assumed that the higher order interaction (tester x component type x configuration) is zero, though strictly this should also be evaluated.

Considering first the interaction of tester with component type, the interaction SSD is given by first subdividing the data into four sets ignoring the circuit configuration variable. Thus each set includes 12 data points so that within each set both remaining variables are held constant.

TABLE 4: ANALYSIS OF VARIANCE CALCULATIONS

| | HEX INVERTER | | QUAD D FLIP-FLOP | | | |
|---|---|---|---|---|---|---|
| | APTTG | APTTF | APTTG | APTTF | MEANS BY TESTER | SSD BY TESTER |
| TESTER I | 2.301 | 3.255 | 3.255 | 3.342 | 3.0140926 | 0.00833 |
| | 2.301 | 2.903 | 2.845 | 3.146 | | |
| | 2.301 | 2.845 | 2.699 | 3.342 | | |
| | 2.602 | 3.255 | 3.255 | 3.342 | | |
| | 2.778 | 3.146 | 3.342 | 3.477 | | |
| | 2.699 | 3.146 | 3.415 | 3.342 | | |
| TESTER II | 2.602 | 3.477 | 2.699 | 3.663 | 3.05134976 | 0.00833 |
| | 2.602 | 3.342 | 2.845 | 3.663 | | |
| | 2.602 | 3.342 | 3.531 | 3.663 | | |
| | 2.477 | 3.146 | 2.477 | 3.342 | | |
| | 2.602 | 2.954 | 2.602 | 3.477 | | |
| | 2.602 | 2.954 | 2.903 | 3.663 | | |
| MEANS BY COMPONENT TYPE | 2.84322 | | 3.22222 | | | |
| MEANS BY CIRCUIT CONFIGURATION | APTTG = 2.7641 | | | | | |
| | APTTF = 3.3013 | | | | | |
| SSD BY COMPONENT TYPE | 0.8620 | | 0.8620 | | | |
| SSD BY CIRCUIT CONFIGURATION | 1.7312 | | | | | |
| | 1.7312 | | | | | |

OVERALL MEAN = 3.0327

TOTAL SSD = 7.6643

Next, the mean of each set is evaluated. The interaction SSD is found by adding the overall mean to each set mean and subtracting the particular tester and component type means which the set is contained in. The (four) results are each then squared and multiplied by the number in each set (which is 12, here). Finally, the four values obtained are summed to give the interaction SSD (for tester x component).

For example, the first term will be given by:

$$\left\{\begin{matrix}\text{Set mean} + \text{Overall} \\ \text{Mean}\end{matrix} - \begin{matrix}\text{Tester I} \\ \text{Mean}\end{matrix} - \begin{matrix}\text{Hex Inverter} \\ \text{Mean}\end{matrix}\right\}^2 \times \text{No. in set}$$

$$\left\{(2.792 + 3.033) - 3.014 - 2.843\right\}^2 \times 12$$

$$\simeq 0.0109$$

and its addition to the three similar terms gives the tester x component interaction as 0.0436.

This is then repeated similarly for the tester x configuration type interaction and the configuration x component type interaction, ignoring in turn the third variable. Full results are in Table 5.

In fact, the only interaction of engineering interest is the tester x configuration; however, it is necessary to statistically evaluate the other two also.

(v) ***Purpose: Calculate the SSD between testers, between configurations and between components.*** The various SSD calculations are now summed to evaluate the effect of each variable. For example, the tester variable is evaluated from $SSD_{\text{Tester I}} + SSD_{\text{Tester II}} = 0.00833 + 0.00833 = 0.01666$.

This quantity is referred to as the "between testers" sum of squared deviations.

The assessment of all the other variables may be similarly constructed.

(vi) *Purpose: Set up an "ANOVA" table.* It is customary to summarize the results, along with some further calculations, in an analysis of variance table. This is done in Table 5.

*Purpose: Calculate the residual SSD.* The residual SSD (found by subtraction) is a measure of experimental error (or "noise").

TABLE 5: ANOVA TABLE

| Source of Variation | SSD | d.f. | MS | F | ? |
|---|---|---|---|---|---|
| Between testers | 0.0166576 | 1 | 0.01666 | 0.323 | Not significant |
| Between components | 1.7237228 | 1 | 1.72372 | 33.372 | Significant |
| Between configurations | 3.462342 | 1 | 3.46234 | 67.033 | Significant |
| INTERACTIONS: | | | | | |
| Tester x configuration | 0.23973 | 1 | 0.23973 | 4.641 | Significant |
| Tester x component | 0.04364 | 1 | 0.04364 | 0.845 | Not significant |
| Component x configuration | 0.060509 | 1 | 0.06051 | 1.172 | Not significant |
| Residuals | 2.11770558 | 41 | 0.05165 | - | - |
| Total | 7.6643 | 47 | | | |

*Purpose: Derive the degrees of freedom for each source of variation.* The SSD column gives the results already explained. "d.f." stands for degrees of freedom, which is given by one less than the number of conditions within a variable; for example, there are two testers, and consequently just one degree of freedom for the between testers SSD. The interaction degrees of freedom are found simply by multiplying the degrees of freedom for the variables being considered. The total degrees of freedom are one less than the number of data points, i.e., 48 - 1 = 47. Again, residual d.f. are found by subtraction.

*Purpose: Calculate the mean squares.* The next column is the mean square (MS) given by MS = SSD/d.f. This essentially standardizes each SSD. Here, all d.f. are 1, but this is not generally so.

*Purpose: Assess the effect of each variable(s) using F.* The penultimate column is the F-statistic which is found by MS/(Residual MS). F is analogous to a signal-to-noise ratio in electronics and actually compares the effect of each variable in turn to the noise in the data. The theoretical value of F has been tabulated (see Chapter 12). This table is consulted by entering at the degrees of freedom for the variable and the residuals. So, in this example, the degrees of freedom consulted in the table are 1 and 41. The level of significance has been previously set at 5%. The critical value of F is therefore found to be 4.08, and, if this value is exceeded, the indication is that that variable <u>does</u> affect failure voltage. (Statisticians put it more precisely by saying that the original null hypothesis (that the variable has no effect) is rejected at the 5% level.)

*Purpose: Determine statistically significant variables and interactions.* The statistically significant variables are indicated in the final column headed "?."

Thus, in this example, with respect to the stated objectives, the statistical conclusions are:

- The two microcircuits exhibit different failure voltages
- The two testers did not obtain significantly different results
- The failure voltages under the APTTG configuration were significantly different from those obtained under APTTF
- There was a significant interaction between the testers and the configurations

These conclusions deserve further discussion. Clearly it is of little use establishing statistical significance if there is no physical explanation. This ties in with the concept of risk: we do have a (1 in 20) risk of drawing an incorrect conclusion in each case.

The microcircuits would be fully expected to exhibit different failure voltages since their construction was entirely different. It had been hoped that the testers would not find different results; the fact that indeed they did not means that their results can be pooled and indicates that the tests were well-specified and, hence, carried out (in general) under very similar conditions.

The difference between APTTG and APTTF results was also fully expected since the floating configuration is much less susceptible to ESD.

The most interesting result was the tester x configuration interaction. Careful inspection suggests this is due to Tester II's APTTF circuit, and it was later revealed that Tester II had used an ionizer prior to measurements. (We later established that the other tester had not done this.) The combined effect of an ionizer with the nature of the APTTF circuit would be expected to produce an interaction, for physical reasons. This gives an indication of the power of analysis of variance, which is more readily appreciated where there are more variables and conditions than were involved in this simple example.

## How Long Does it Take?

The example used was completed by hand within a couple of hours using a calculator. Computer programs are available, but it is the author's view that the extra time taken manually is more than compensated for by the degree of understanding of the data yielded by working with it. For very large data arrays, computer solution may become more useful. There are other time-saving methods for manual work which have been omitted here for simplicity.

## Numerical Solution

Notice that the analysis has so far identified the variables which affect failure voltage but has not quantified their effect. Thus, for example, we do not yet know whether the hex inverter or the quad D flip flop exhibited the higher voltage and by how much. (Graphically, the answer seems to be the latter, but the "how much" is less obvious.) Numerical solution of the mathematical model is straightforward but is postponed until Chapter 11.

## What if the Model Assumptions are Violated?

At least as often as not, the assumptions of normal error, equal variance and balanced samples do not hold. The obvious options of turning in (a) an invalid analysis or (b) nothing might be a little embarassing. Fortunately, we have nonparametric analysis of variance which assumes only reasonably similar and symmetric dispersion. There are different methods for different data structures. For a two-way analysis along the lines of what was just done, there is Friedman's method (Reference 7). For a simpler case of only one variable, there is the Kruskal-Wallis method.

## The Kruskal-Wallis One-Way Method (Example 2)

The following data were collected on the same equipment used in five different types of aircraft as failures per $10^6$ running hours. Clearly, the sample is unbalanced with unequal samples for each aircraft.

TABLE 6: AIRCRAFT DATA I

| Attack | Bomber | Cargo | Fighter | Trainer |
|---|---|---|---|---|
| 3364 | 1800 | 786 | 2207 | 1165 |
| 3278 | 1488 | 758 | 1457 | 776 |
| 1147 | 1000 | 1136 | 1632 | 940 |
| 1621 | 1107 | 466 | 1727 | 1692 |
| 1229 | 821 | 688 | 1750 | 652 |
| 1854 | | 1207 | 1368 | 655 |
| 1228 | | 622 | 2381 | 653 |
| 709 | | 599 | 1000 | 621 |
| 1043 | | 709 | 1562 | |
| | | 649 | 1333 | |
| | | 615 | 1474 | |
| | | 579 | | |

The procedure is as follows:

(i) Rank the data smallest to largest as in Table 7 below (with the data arranged as before for clarity); where two or more observations are equal, the ranks are added and divided by the number of equalities.

(ii) Evaluate $H = \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(N+1)$

where

$k$ = no. of groups
$n_j$ = no. in $j^{th}$ group
$N = \Sigma n_j$
$R_j$ = sum of ranks in $j^{th}$ group

It may be shown that H has a $\chi^2$ distribution with k-1 degrees of freedom.

TABLE 7: AIRCRAFT DATA II

| Attack | Bomber | Cargo | Fighter | Trainer |
|---|---|---|---|---|
| 45 | 40 | 16 | 42 | 25 |
| 44 | 33 | 14 | 31 | 15 |
| 24 | 19½ | 23 | 36 | 18 |
| 35 | 22 | 1 | 38 | 37 |
| 28 | 17 | 11 | 39 | 8 |
| 41 | | 26 | 30 | 10 |
| 27 | | 6 | 43 | 9 |
| 12½ | | 3 | 19½ | 5 |
| 21 | | 12½ | 34 | |
| | | 7 | 29 | |
| | | 4 | 32 | |
| | | 2 | | |

Hence,

$k = 5$ $\quad R_1 = 277.5$

$n_1 = 9$ $\quad R_2 = 131.5$

$n_2 = 5$ $\quad R_3 = 125.5$

$n_3 = 12$ $\quad R_4 = 373.5$

$n_4 = 11$ $\quad R_5 = 127$

$n_5 = 8$

$N = 45$

So

$$H = \frac{12}{45(46)} \left( \frac{277.5^2}{9} + \frac{131.5^2}{5} + \frac{125.5^2}{12} + \frac{373.5^2}{11} + \frac{127^2}{8} \right) - 3\,(45 + 1)$$

$$= 162.47 - 138$$

$$= 24.47$$

Since H has a $\chi^2$ distribution, we use tables to find the 5% critical value (for K-1 = 4 degrees of freedom) as 9.49. The value 24.47 is found to be significant (exceeds 9.49) and hence we reject the null hypothesis that aircraft type does not affect failure rate.

## Philosophy of Multiple Comparison Tests

One might wonder why it is not possible to just run a t test or similar between two groups of aircraft data at a time and keep going until all differences are established. This would give an incorrect answer because the significance levels would be wrong. Repeated tests compound the risk of error.

## Review

Complex data structures (which statisticians term "factorial data") can be reduced and analyzed by analysis of variance which uses the F

statistic to make decisions by analogy with a signal-to-noise ratio. The corollary to analysis of variance is experimental design: how to get the maximum out of an experiment for minimal resources and a minimum amount of data. Randomisation is also important; that is to say that as long as each factor is randomly sampled among all other factors, a valid analysis may be performed. It is preferable to start thinking of an experimental design before data collection rather than thinking of an ANOVA after collecting data. In fact, Example 1 was an engineer/statistician collaboration. It is appreciated that this is often difficult; in practice one may not be able to use this technique, either because of a lack of data or because the normality assumption is not justified. In such cases nonparametric methods may be preferable. Numerical solution of the parametric method is discussed in Chapter 11.

# CHAPTER 3

## Confidence Intervals

CHAPTER 3

# Confidence Intervals

## Summary

This note defines and discusses statistical confidence interval estimation and looks at some methods which may be useful in reliability studies.

## Introduction and Definition

A point estimate of some parameter is not meaningful without some measure of its possible error. Such a measure, which is often used, is the confidence interval. Different samples yield different intervals; some of these intervals will contain the true (unknown) parameter and some will not. It is possible to define a $100(1 - \alpha)\%$ interval such that $100(1 - \alpha)\%$ of the intervals (of which ours is just one) will contain the parameter. Usually we compute two-sided central confidence intervals which lie symmetrically about the parameter with equal risks of the parameter being excluded on either side. It is also possible to compute one-sided intervals which just give an upper or lower confidence limit and non-central intervals which have unequal risks in each tail. The confidence level, $1 - \alpha$, is expressed between 0 and 1 although in practice it is often expressed as a percentage. Confidence intervals may be constructed on any parameter, e.g., the mean, median, variance, 90th percentile, regression slope, etc.

Note that the concept of a confidence interval assumes the existence of some true parameter value, which is unknown but which we are trying to estimate from data. In classical statistics, the true parameter value is assumed to be fixed and the probabilistic

interpretation of a confidence interval is defined above. Bayesian intervals are conceptually different to the classical approach and are outside the scope of this note, which restricts itself to classical and nonparametric methods. The interested reader may pursue Bayesian methods in References 8 and 9 although some preliminaries are given in Chapter 7.

At the Drop of a Hat - An Analogy

Some readers will prefer a non-mathematical explanation of confidence intervals, which also brings out the subtlety of their probabilistic interpretation.

Consider an orange lying on a desk. I might stand at the opposite side of the office and attempt to throw my hat so that it lands on the orange, obscuring it from view. If I find I can cover the orange 80% of the time I throw the hat, then my hat (when thrown) defines an 80% confidence interval for the position of the orange. Notice that 80% is not the probability of the orange being inside the "interval" (either it is, or it is not); the 80% probability relates to my throwing the hat.

Now suppose my office colleague (who has a much smaller head) throws his hat at the same orange. Since his hat is smaller he only covers the orange 60% of the time (assuming our throwing skills are equal). He therefore defines a 60% interval on the position of the orange.

If we repeat the procedure with a sombrero (of 3 feet diameter) and find the orange is covered 99% of the time, then clearly this is by analogy a 99% confidence interval for the orange's position. It may seem that this is the "best" option since the sombrero is virtually certain to cover the orange. However, the problem lies in our

uncertainty within the "interval." If I take a hatpin and try to pierce the orange through the sombrero, I have a good chance (spatially) of missing the orange. Following the same procedure with my colleague's hat, it is almost certain that the orange will be pierced with the pin; however, his hat will only cover the orange 6 times out of 10 and we will not know when it covers the orange and when it doesn't. My own hat represents a compromise between the two extremes.

Though the analogy misses the distributional constraints of confidence interval estimation, it does illustrate the general concept. Reverting to statistics, the orange is some parameter we're trying to estimate and the hat is a confidence interval constructed from data. For the sake of a little more mathematical rigor (as well as my colleague's hat) the basic theory is now developed.

## The Concept of a Confidence Interval for the Normal Distribution

The concepts that follow are developed slowly from basic familiar ideas to avoid misinterpretation.

If x is a random variable and is known to be normally distributed, then its probability density function (p.d.f.) may be sketched as below:

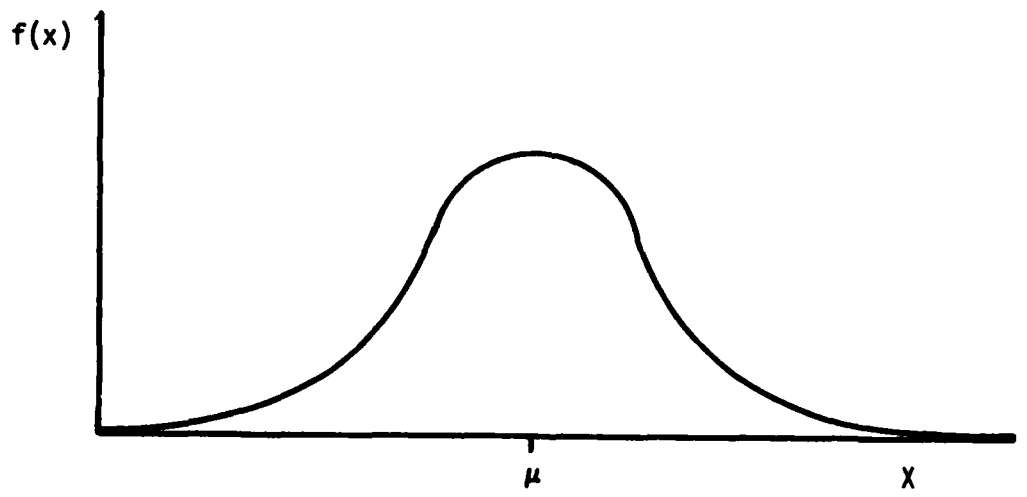



FIGURE 8: THE NORMAL DISTRIBUTION

If the true value (of whatever is being measured) is $\mu$ then the dispersion indicated in the sketch is attributable to a variety of sources including measurement error, experimental error, fluctuating conditions and environments. The crux of modern statistics is that the overall effect of these errors may be grouped under a general title, "sampling error," and assessed empirically using the well-known measure termed the variance, in this case having a true value of $\sigma^2$, where $\sigma$ is the standard deviation. (Notice that $\mu$ is analogous to the center of gravity in statics, $\sigma^2$ is analogous to moment of inertia in dynamics. $\sigma$ is crudely analagous to radius of gyration.)

An actual sample of x (which might be the number of hours of use of an equipment per year) would exhibit a mean at $\bar{x}_1$ and a variance of $S_1^2$. $\bar{x}_1$ and $S_1^2$ may or may not be equal to $\mu$ and $\sigma^2$. Since generally we do not know $\mu$ or $\sigma^2$, we estimate them by $\bar{x}_1$ and $S_1^2$.

Now if a series of m samples were taken, a number of estimates of $\mu$ and $\sigma^2$ would result as $\bar{x}_2$, $\bar{x}_3$ ... $\bar{x}_m$ and $S_1^2$, $S_2^2$, $S_3^2$ ... $S_m^2$. A plot of the m values of $\bar{x}$ would tend to be grouped with much less dispersion than the original data. The values of $\bar{x}$ are in fact m observations from the distribution of the mean which may be sketched (with the original distribution of x superimposed by the dotted line) thus:

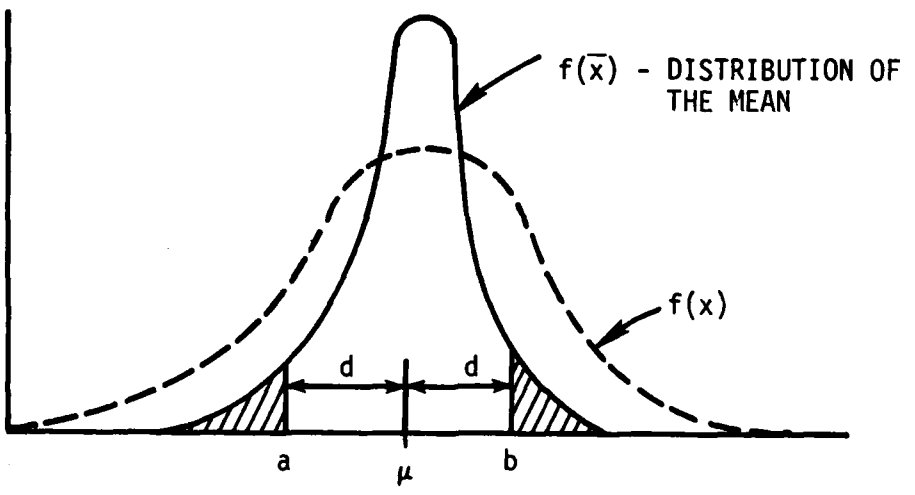



FIGURE 9: THE DISTRIBUTION OF THE MEAN OF A NORMAL

The distribution of the mean is in fact also normal with (theoretical) mean $\mu$ and variance $\sigma^2/n$, where n is the number of observations per sample; hence, the larger the sample size n, the less variability in the distribution of the mean. The standard deviation is now $\sigma/\sqrt{n}$ which is termed the standard error.

It may now be seen intuitively that the distribution of the mean might be used to place some probabilistic interpretation on the likely error in the use of an $\bar{x}$ as an estimator of $\mu$. In fact, we'd expect that, if the shaded areas in the sketch each represent 5% of the distribution of the mean, then 90% of the m values of $\bar{x}$ would lie between a and b. So conversely it is expected that $\bar{x} \pm d$ would include $\mu$ for only 90% of our m samples, with d defined as indicated on the sketch. However, it should be made clear that it is the interval which is subject to statistical fluctuation, and not $\mu$. (Recall the hat and the orange analogy.) Hence the 90% used in developing the interval $\bar{x} \pm d$ is termed confidence and is not synonymous with probability. In fact, confidence may be defined at any level; 90% was used for convenience in developing the example. Thus, $\bar{x} \pm d$ is a 90% confidence interval for $\mu$.

d may be evaluated mathematically by introducing the transformation

$$z = \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}} \tag{4}$$

Then z is distributed normally with mean 0, variance 1. (z is used for convenience since its distribution percentiles are tabulated).

A value, c, may be defined so that

$$\text{Prob.} \left[ -c \leq \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}} \leq c \right] = 1 - \alpha \tag{5}$$

where $1 - \alpha$ is the confidence. Rearranging,

$$\text{Prob.} \left[ \bar{x} - c\sqrt{\sigma^2/n} \leq \mu \leq \bar{x} + c\sqrt{\sigma^2/n} \right] = 1 - \alpha \tag{6}$$

So a $1 - \alpha$ central confidence interval on $\mu$ is given by:

$$\bar{x} \pm c\sqrt{\sigma^2/n}$$

where c is found from tables (Reference 3) of the standardized normal distribution percentiles. So, for example, if $1 - \alpha$ is chosen as 0.9 (i.e., 90%) c is found to be 1.6.

In practice, $\sigma^2$ will usually have to be estimated from the sample.

Note that, for $1 - \alpha = 0.95$ (i.e., 95%), c is found to be 1.96 giving rise to the commonly used rule of thumb of taking $\pm$ 2 standard errors from the sample mean for an approximate interval.

Example

In a recent study, IIT Research Institute evaluated the performance of a variety of the models given in MIL-HDBK-217C (Reference 10). Though data on dozens of components were collected, only one generic type is considered for this example.

Field failure rate data on a set of 62 CMOS devices were compared to the failure rates predicted by MIL-HDBK-217. If the field failure rate is $\lambda_o$ and the predicted failure rate $\lambda_p$, then it was found that $\log_{10}(\lambda_o/\lambda_p)$ was distributed (to a reasonable approximation) normally with sample mean -0.022 and sample standard deviation 0.413. The extent to which the normal fit is reasonable was evaluated by the Kolmogorov-Smirnov statistic (Chapter 4). The reason for the fit is discussed in the study report (Reference 11). The data and the probability plot are sketched on the following pages:

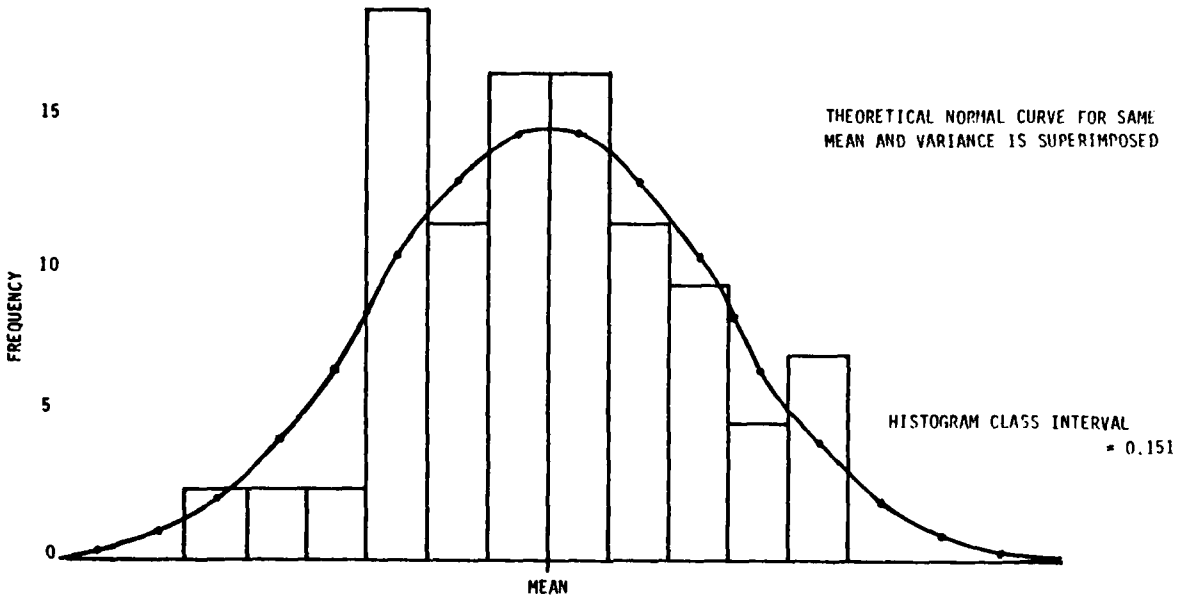



FIGURE 10: DISTRIBUTION OF $\text{LOG}_{10}\ \lambda_o/\lambda_p$

Now, if the prediction model were performing perfectly, than $\lambda_o$ would equal $\lambda_p$; hence $\log_{10}(\lambda_o/\lambda_p) = \log_{10}(1) = 0$. Our sample would be expected to have some dispersion about the perfect value (0); the sample mean was -0.022, which seems small but it is not clear what is acceptable.

So we might construct a 95% confidence interval on the sample mean using formula (6), which gives the interval as

$$-0.022 \pm 1.96 \times \frac{0.413}{\sqrt{62}}$$

i.e., -0.125 to 0.081

Thus we have (at least) 95% confidence that the true performance of the model (as measured by $\log_{10}(\lambda_o/\lambda_p)$) is between -0.125 and 0.081. Since this interval includes the ideal value, things are as they should be.

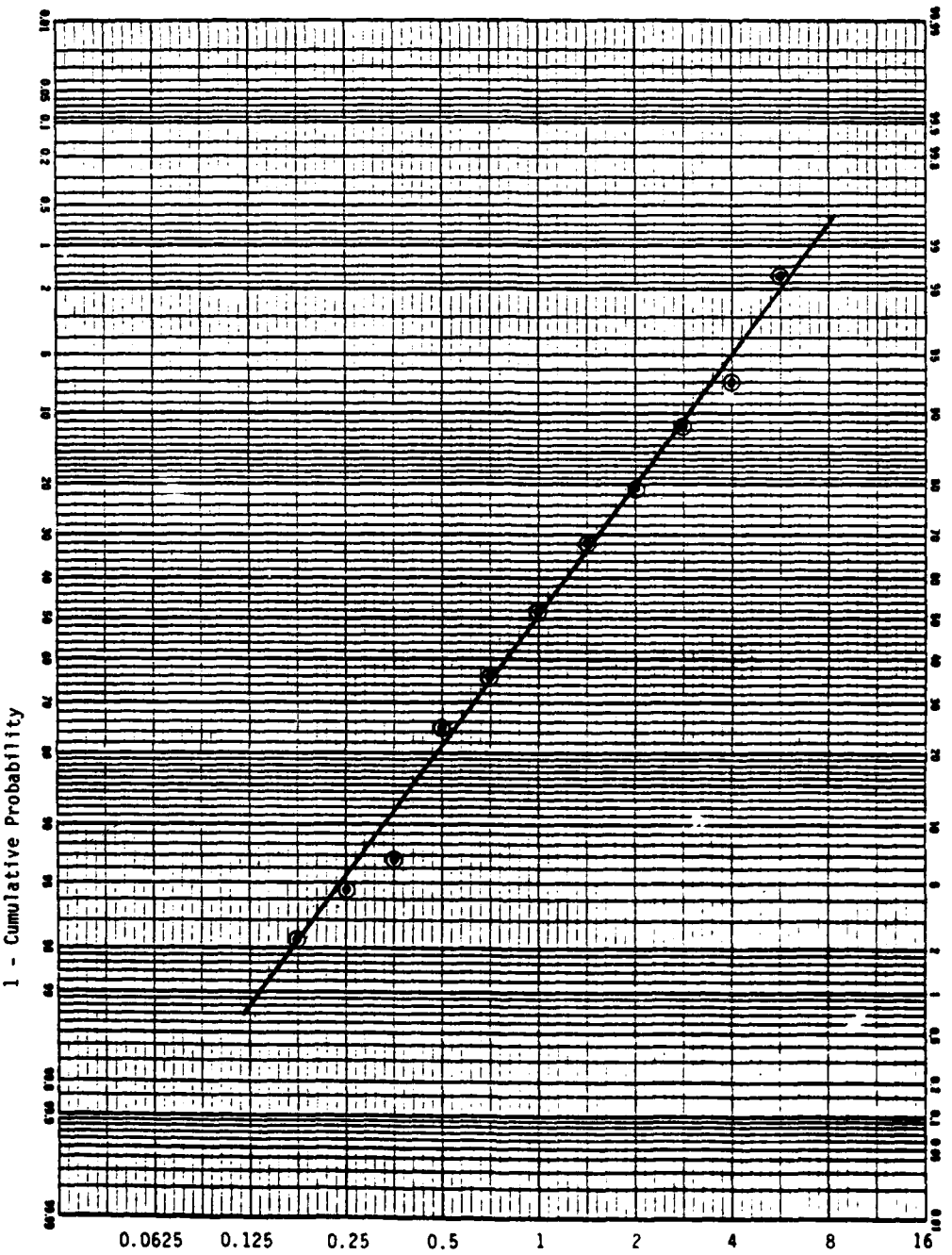



FIGURE 11: PROBABILITY PLOT (NORMAL) FOR DATA

Two points should be noted. First, it is more conventional to set up a hypothesis test for this type of problem, hypothesis tests being a direct corollary to confidence interval estimation. Generally, confidence intervals are used where we don't know the ideal position of the parameter. This example was chosen to illustrate more vividly. Second, we have only measured performances with respect to position; a complete model critique would include an investigation as to the variability accounted for, probably by some variation on the multiple coefficient of determination (Chapter 9). Schafer et al. give a nice treatment (Reference 12) in evaluating growth models.

The principle of deriving a confidence interval on the mean of a distribution applies regardless of its form. However, as the following paragraphs show, the method is not usually simply a matter of adding (subtracting) a certain number of standard errors from the mean.

Confidence Intervals for the Mean of an Exponential

Many distributions encountered in reliability work are of a form other than normal. Chapter 6 provides greater detail on choice of distribution, but it is found that for component failures which are random in time, the distribution of time-to-failure (TTF) is exponential. For different reasons the distribution of time between system failures is also found to be exponential, typically. So a means of establishing confidence intervals for the mean of an exponential distribution would be useful.

Suppose the distribution on the following page is that of TTF for a certain part. Then the distribution of the mean TTF is shown as the dotted line, and it is readily shown that this distribution is gamma, which is a little inconvenient for practical computations. However, fortunately we can make use of a relation to the chi-square to provide a simple formula for the upper and lower limits of the interval. Suppose

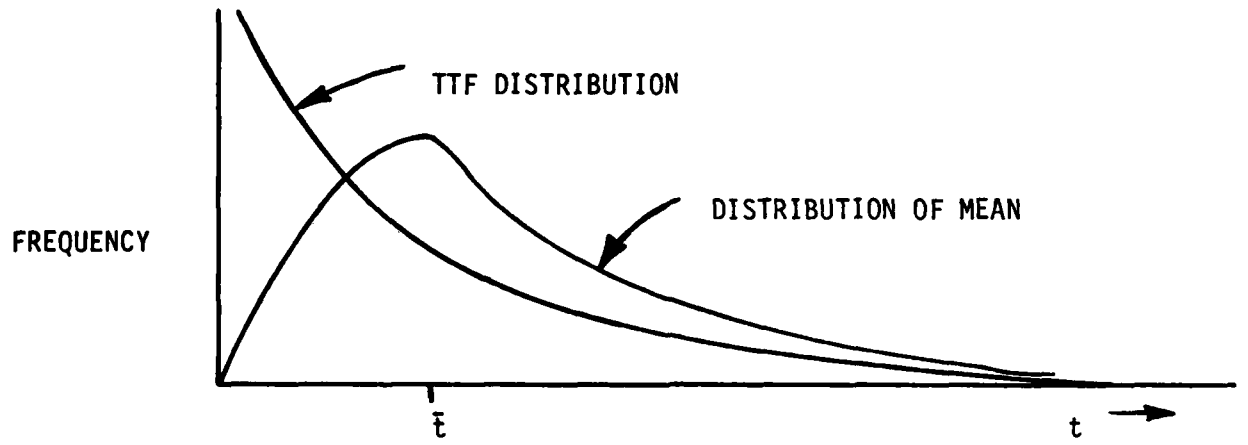



FIGURE 12: THE EXPONENTIAL DISTRIBUTION

r parts fail in a total of T part hours. Then the mean TTF (MTTF) estimate is given by T/r and the confidence interval is defined by:

$$\text{Upper limit} = \frac{2T}{\chi^2_{2r,\ 1-\alpha/2}} \tag{7}$$

$$\text{Lower limit} = \frac{2T}{\chi^2_{2r,\ \alpha/2}} \tag{8}$$

where $(1 - \alpha)$ is the confidence required (which may be expressed as 100 $(1 - \alpha)\%$). $\chi^2$ is the chi-square distribution evaluated from tables with 2r degrees of freedom at the stated probability of $1 - \alpha/2$ or $\alpha/2$.

T is the sum of the individual TTFs, i.e., the total part hours.

If the sample is from a life test or similar, where the test was terminated before completion, then an estimate of the MTTF and its confidence interval is still possible though we must also include the test time of those parts which have not yet failed. However, there is the danger of obtaining biased estimates. There are two types of test termination (prior to failure of all parts) which are termed "time truncation" and "failure truncation" otherwise known as Types I and II

censoring. The former involves terminating the test at a certain time and evaluating the number of parts which failed; the latter involves termination after a certain number of failures and estimation of the test time sustained.

Often, data are expressed in terms of failure rate observed in a given time interval (e.g., $10^6$ hours). Such data form a Poisson process with mean $\lambda\tau$, where $\lambda$ is the observed failure rate and $\tau$ the time interval. $\lambda$ is estimated by:

$$\hat{\lambda} = \tau r/T$$

MIL-HDBK-217 typically takes $\tau = 10^6$ hours and failure rates are therefore expressed as "parts per million hours" (p.p.m.h.). Confidence intervals on the failure rate are given by:

$$\text{Upper limit} = (\chi^2_{2(r+1),\ \alpha/2})/(2T/\tau) \tag{9}$$

$$\text{Lower limit} = (\chi^2_{2r,\ 1-\alpha/2})/(2T/\tau) \tag{10}$$

with the symbols as before.

Note that for the MTTF confidence interval the random variable is time (a continuous variable) whereas for the failure rate confidence interval the random variable is number of failures per time interval (which is a discrete variable) and this affects the degrees of freedom. TTF is the reciprocal of failure rate (if $\tau = 1$) although note that the former is distributed as exponential and the latter as Poisson.

Another Example

Six failures of a particular commercial grade TTL microcircuit operating in a controlled environment were observed in $24.24 \times 10^6$ part

hours. We require a 90% confidence interval as the mean.

Choose $\tau = 10^6$ hours; then the failure rate estimate is given as:

$$\hat{\lambda} = \frac{10^6 \times 6}{24.24 \times 10^6}$$

$$= 0.248 \text{ p.p.m.h.}$$

A 90% confidence interval as $\hat{\lambda}$ is found from (9) and (10) by setting $(1 - \alpha) = 0.9$ and since $T = 24.24 \times 10^6$ hours, and $r = 6$, the upper limit $\lambda^*$ is given by $\lambda^* = 23.685/(2 \times 24.24 \times 10^6)/10^6$ given by:

$$\lambda^* = \frac{23.685}{(2 \times 24.24 \times 10^6)/10^6}$$

$$= 0.489 \text{ p.p.m.h.}$$

The lower limit $\lambda_*$ is given (from (10)) by:

$$\lambda_* = \frac{5.226}{(2 \times 24.24 \times 10^6)/10^6}$$

$$= 0.108 \text{ p.p.m.h.}$$

Thus, the data yield a best estimate of the failure rate as 0.248 p.p.m.h. and a 90% central confidence interval as $\hat{\lambda}$ is given by (0.108, 0.489) p.p.m.h.

## Confidence Intervals for Parameters of the Weibull Distribution

Confidence intervals may be constructed, though not always very easily, for the parameters of the Weibull distribution. Such methods are rarely analytical; they are usually empirical or Monte Carlo simulated. In view of their complexity, it is advisable to refer to texts on the subject. Some easy-to-follow methods are given in Reference 13 and some more advanced treatments in Reference 8.

## Nonparametric Confidence Intervals on the Mean (For Complete Samples)

Nonparametric methods may be used where there is not much data, where it is measured with uncertainty, or where the distribution is unknown. Such methods consider the probability of certain sequences of ranks occurring, and often use the median as an estimate of the center of the distribution. A confidence interval on a mean may be obtained from uncensored data using the relation

$$\text{Prob}\left[Z_{(i)} < \mu \leq Z_{(i+1)}\right] = \binom{N}{i}\left(\frac{1}{2^N}\right) \tag{11}$$

where $Z_{(i)}$, $Z_{(i+1)}$ are the i and $i+1^{th}$ ranked data points (ranked smallest to largest). N is the number of data points. $\binom{N}{i}$ is the binomial coefficient, e.g., if observations are 2, 3, 4, 5, 6, and 7.

$$\text{Prob}\left[Z_{(3)} < \mu \leq Z_{(4)}\right] = \binom{6}{3}(1/2^6) = 0.3125 \tag{12}$$

which is simply saying that the 3rd and 4th (ordered) observations define 31.25% confidence interval on the median. Clearly this method is limited and one can get closer to the usual confidence levels by considering

$$\text{Prob}\left[\mu \leq Z_{(i)}\right]$$
$$\text{Prob}\left[\mu > Z_{(j)}\right]$$
$$i > j$$

This is done by evaluating

$$\sum_{u=o}^{i \text{ or } j} \binom{N}{u}\left(1/2^N\right) = \text{confidence level on i or } j^{th} \text{ ordered point}$$

(choosing i and/or j to give the required level of significance).

Some examples of the first few upper confidence limits, for N = 2 to N = 10 are given below, derived from the above method. Note that nonparametric probabilities are never as "neat" as parametric. We have to take what we can get.

TABLE 8: NONPARAMETRIC CONFIDENCE LIMITS

| N | Confidence Level | Upper Confidence Limit |
|---|---|---|
| 2 | 75% | $Z_{(2)}$ |
| 3 | 87.5% | $Z_{(3)}$ |
| 4 | 93.75% | $Z_{(4)}$ |
| 5 | 96.87% | $Z_{(5)}$ |
| 6 | 98.4% | $Z_{(6)}$ |
| 7 | 99.2% | $Z_{(7)}$ |
| 8 | 99.6% | $Z_{(8)}$ |
| 9 | 98.1% | $Z_{(8)}$ |
| 10 | 98.9% | $Z_{(9)}$ |

Lower limits may be readily derived by the same formula and hence central confidence intervals may be found. Note that the method only applies to complete samples where all parts have failed, i.e., there is no censoring.

A fuller explanation of nonparametric confidence intervals is given in Reference 14.

## Concluding Remarks

As explained in the introduction to this note, a confidence interval may be established on any parameter. Regression coefficients and slopes of fitted lines themselves have confidence intervals which

may be derived in each case by considering the relevant distribution. Confidence may also be given if necessary in more than one dimension; for example, a confidence ellipse results on the mean of a bivariate normal distribution (for which the hat and orange example is a truer analogy).

For complex problems of interval estimation, it may be worth considering a computationally intensive algorithm termed the jacknife. This method produces results remarkably close to parametric methods, based on repeated resampling within the original sample. The method is not covered in this note, but full details are given in References 28 and 31.

# CHAPTER 4

## Goodness of Fit Tests

CHAPTER 4

# Goodness of Fit Tests

## Summary

Reliability data is often fitted by statistical distributions. Without some statistical measure, it is impossible to form an objective assessment of how well the distribution fits the data. This note describes two of the more popular techniques, the Chi-square test and the Kolmogorov-Smirnov test, both methods being nonparametric.

## The Chi-Square ($\chi^2$) Test

However well a distribution fits a particular set of data, we would expect some "noise", or experimental errors. A hypothetical example by histogram is given below.

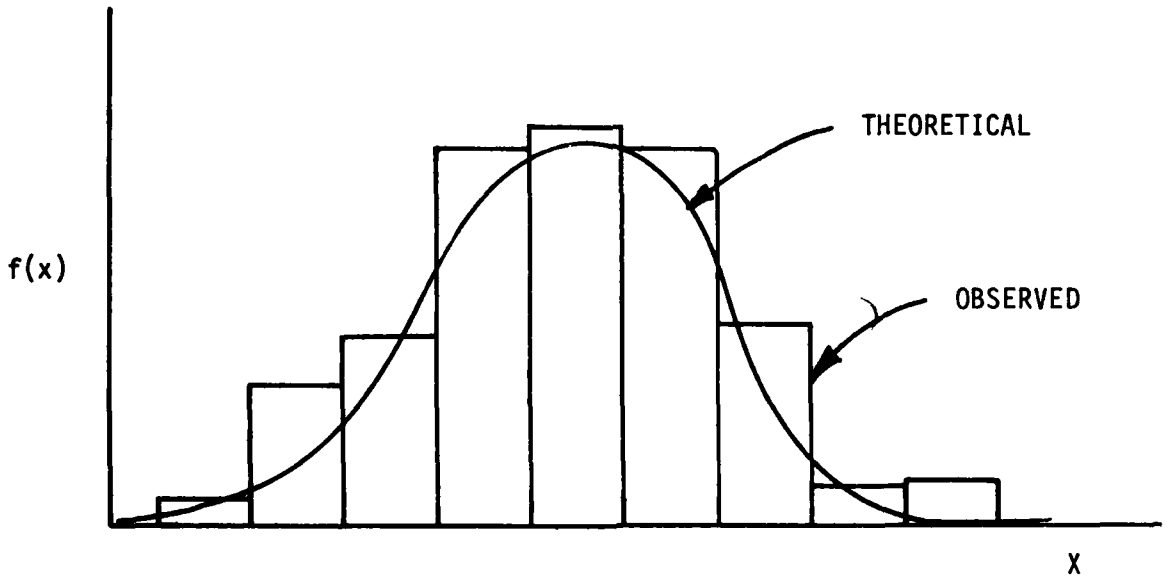



FIGURE 13: GOODNESS OF FIT - HISTOGRAM AGAINST p.d.f.

The Chi-square ($\chi^2$) test is derived probabilistically from the number of observations expected to fall into each interval (if the correct distribution was chosen). Hence the amount of acceptable deviation from perfect fit is determinable. In this way it is possible to decide whether or not the data came from some hypothesized distribution. The $\chi^2$ test is a discrete method since it is defined on data collected in the form of a histogram. (Hence it's rather like putting an A-D converter in between the raw data and the decision process.) For a truly analog test, try Kolmogorov-Smirnov or apply Yate's continuity correction (Ref. 5).

The chi-squared statistic is evaluated by subtracting the expected from the observed frequency for each class interval on which the histogram of data is defined, squaring the result, and dividing by the expected frequency. So:

$$\chi^2 = \sum_{i=1}^{k} (O_i - E_i)^2/E_i \qquad (12)$$

where $O_i$ is the observed frequency in the ith class interval. $E_i$ is the expected frequency in the ith class interval, calculated from the hypothesised distribution. k is the total number of classes in the histogram.

If there are less than five observations in any class interval, then the data should be pooled with the adjacent class interval. The expected frequencies are adjusted accordingly. The value of $\chi^2$ is then compared to the critical value, which is given in tables of the chi-squared distribution for the appropriate number of degrees of freedom. In the $\chi^2$ test we lose a degree of freedom for the artificial grouping into k classes (because once all but one of the frequencies are defined, the last is defined by subtraction from the total number of observations). A further p degrees of freedom are lost where p is the

number of parameters (e.g., standard deviation) independently estimated from the data. So, the degrees of freedom is given by k-1-p.

Choosing the histogram class interval can be tricky, particularly if there is not much data. Generally, though, it is chosen so that most intervals have more than five observations. If this is not possible, use a different test such as Kolmogorov-Smirnov. An example is now given.

Example I

Data for component failure times was collected and is summarized below (in terms of x 10,000 hours). They have been ordered smallest to largest for simplicity.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0.02 | 0.45 | 0.52 | 0.60 | 0.99 | 1.35 | 1.78 | 1.83 |
| 2.16 | 2.22 | 2.34 | 2.94 | 3.43 | 4.82 | 5.32 | 5.32 |
| 6.44 | 7.8 | | | | | | |

Suppose we have just done a Weibull plot and find the data appears to be close to exponential. The question now is how close is it? We start by defining the exponential probability density function:

$$f(t) = \frac{1}{\theta} e^{-t/\theta} \qquad (13)$$

Where t is time and $\theta$ is the mean time to failure (MTTF).

Next we need to estimate $\theta$ from the data. Note that ideally we would know $\theta$ in advance, perhaps from previous test results. There is no problem with estimating $\theta$ from the data except that we lose a degree of freedom and this does not really concern us.

From the data, $\hat{\theta} = 2.796 \times 10^4$ hours.

We now choose a class interval of $2 \times 10^4$ hours and group the data into a histogram format as follows; this gives the observed frequencies:

TABLE 9: HISTOGRAM GROUPING OF DATA

| Class of t ($\times 10^4$ hours) | Number in Class (Frequencies) |
|---|---|
| 0 - 2 | 8 |
| 2 - 4 | 5 |
| > 4 | 5 |

A class interval of $2 \times 10^4$ is chosen since it gives two groups of 5 or more, and the final one represents a pool of two classes (4 to 6 and 6 to 8).

We now evaluate the expected frequencies from (2) (or from tables of the cumulative exponential distribution). The cumulative distribution function associated with (13) is

$$F(t) = 1 - e^{-t/\theta} \tag{14}$$

This function is evaluated at each class boundary (e.g., $t = 2 \times 10^4$, $4 \times 10^4$ etc.) and the proportion in each class is then derived by successive subtractions (since F(t) is cumulative). It is simple to then calculate the expected frequencies by multiplying the proportion in each class by the total number of observations (18 in this example). The calculations are tabulated below. The observed frequencies are also given as before, for completeness.

TABLE 10: OBSERVED AND EXPECTED FREQUENCIES

| Class Interval ($\times 10^4$ hours) | Observed Frequency ($O_i$) | t ($\times 10^4$) | F(t) | Expected Proportion in Class Interval | Expected Frequency ($E_i$) |
|---|---|---|---|---|---|
| 0 - 2 | 8 | 2 | 0.51 | 0.51 | 9.2 |
| 2 - 4 | 5 | 4 | 0.76 | 0.25 | 4.5 |
| > 4 | 5 | $\infty$ | 1 | 0.24 | 4.3 |

The $\chi^2$ statistic is then evaluated from (12) as

$$\chi^2 = \sum_{i=1}^{3} (O_i - E_i)^2/E_i$$

$$= \frac{(8 - 9.2)^2}{9.2} + \frac{(5 - 4.5)^2}{4.5} + \frac{(5 - 4.3)^2}{4.3} = 0.326$$

Now k = 3 since there are three classes in the data when arranged into histogram format. p = 1 since only one parameter ($\theta$) was estimated from the data.

So the degrees of freedom = 3 - 1 - 1 = 1

Hence we consult the $\chi^2$ tables with 1 degree of freedom and find that our value does not exceed the critical value. (At 5% significance, the critical $\chi^2$ is 3.841.) Hence we may conclude that our data is not significantly different from the exponential distribution with parameter $\theta = 2.796 \times 10^4$.

Usually k would be larger than 3 but this example was chosen to keep the numbers simple.

Note that the $\chi^2$ test can only be used to compare frequencies as shown here.

## The Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (K-S) goodness-of-fit test performs essentially the same function as the $\chi^2$. It is also a nonparametric test. Instead of quantizing the data first, the K-S test is a continuous method since it looks directly at the observed cumulative distribution function and compares it to the theoretical (or expected).

Any cumulative distribution function $F_O(t)$ may be evaluated from data at some point t* by

$$F_O(t^*) = \frac{\text{Number of components failing by } t^*}{\text{Total number of components on test} + 1}$$

The denominator includes the term "+1" to reduce the bias. There are other methods for doing this, discussed in Chapter 6.

The theoretical cumulative density function $F_E(t)$ may be evaluated by integrating the probability density function f(t).

If the observed expected cumulative distribution functions are $F_O(x)$ and $F_E(x)$, then

$$D = \max \left| F_O(x) - F_E(x) \right| \tag{15}$$

is the K-S statistic,* i.e., it is the largest deviation of observed from expected, on the cumulative curve. D is then compared to tables of critical values (Appendix 2) and if it exceeds the critical value (at some predetermined level of significance) then it is concluded that the observations do not fit the theoretical distribution chosen.

*Actually this is strictly the K-S one-sample statistic. A two-sample procedure exists also for comparing two samples to each other.

Example II

The same data as before is used. The observed and expected cumulative distributions are evaluated at all possible values of t. The observed F(t) is evaluated at the ith observation by $F_O(t) = i/(n + 1)$ (where n is the total number of observations). The expected is evaluated from Equation (14).

TABLE 11: OBSERVED AND EXPECTED CUMULATIVE FREQUENCIES

| t ($\times 10^4$) | $F_O(t)$ | $F_E(t)$ | $F_O(t) - F_E(t)$ |
|---|---|---|---|
| 0.02 | 0.053 | 0.007 | 0.046 |
| 0.45 | 0.105 | 0.149 | -0.044 |
| 0.52 | 0.158 | 0.170 | -0.012 |
| 0.60 | 0.211 | 0.193 | 0.018 |
| 0.99 | 0.263 | 0.298 | -0.035 |
| 1.35 | 0.316 | 0.383 | -0.067 |
| 1.78 | 0.368 | 0.471 | -0.103 |
| 1.83 | 0.421 | 0.480 | -0.059 |
| 2.16 | 0.474 | 0.538 | -0.064 |
| 2.22 | 0.526 | 0.548 | -0.022 |
| 2.34 | 0.579 | 0.567 | 0.012 |
| 2.94 | 0.632 | 0.651 | -0.019 |
| 3.43 | 0.684 | 0.707 | -0.023 |
| 4.82 | 0.737 | 0.822 | -0.085 |
| 5.32 | 0.789 | 0.851 | -0.062 |
| 5.32 | 0.842 | 0.851 | -0.009 |
| 6.44 | 0.895 | 0.900 | -0.005 |
| 7.8 | 0.947 | 0.939 | 0.008 |

D is found as the modulus of the maximum deviation of observed from expected, so D = 0.103. The critical value for n = 18, is 0.309 (at 5% significance) or 0.244 (at 20% significance). Hence our value is not significant and we conclude that the observations are not significantly different from the exponential model with $\theta$ as before. (This result is the same as for the Chi-square test).

Comparison of the Two Tests

In recent years the K-S test has become as popular as the $X^2$ test (if not more so). Note that it is very easy to do the K-S test directly from a Weibull plot (see Chapter 6) simply by reading off the maximum % deviation of the observations from the fitted line, dividing by 100, and comparing to the critical values.

Studies on the power of the tests have found the K-S test to be superior; however, in caution it is worth noting that the power of the K-S test is reduced if the parameter has to be estimated from the sample (which it usually does).

The Chi-square test cannot be applied to extremely small samples because of the grouping procedure, whereas the K-S test is unrestricted in this respect.

The relative merits of the tests are tabulated below:

| | Kolmogorov-Smirnov | Chi-Square |
|---|---|---|
| Power | Superior | Good |
| Sample Size | Unrestricted | Less effective for small samples |
| Implementation | Direct. May be read approximately from a Weibull plot | Requires judgement to form grouping intervals. |

## Review

The goodness-of-fit of a given set of data to some hypothesized distribution is measurable by either the Chi-square or the Kolmogorov-Smirnov test. The choice of distribution hypothesis is generally made on one or both of:

- o knowledge of underlying physical principles involved
- o observed data structure

The two methods are compared for power, restrictions, and implementation.

# CHAPTER 5

## Sampling Inspection and Statistical Q.C.

CHAPTER 5

# Sampling Inspection and Statistical Q.C.

## Summary

The need for sampling procedures in industry is introduced and extended to commonly used plans such as Acceptable Quality Level (AQL) and Lot Tolerance Percent Defective (LTPD). Detailed examples of the use of LTPD plans and their implementation in MIL-M-38510, MIL-S-19500 (and associated test specifications) are given. The relation between the various plans is defined using the operating characteristic curve concept. The development of these ideas into more complex methods such as sequential plans and design of life tests is discussed.

## Introduction

Many areas of industry call for quality control. Usually it is not cost-effective or physically possible to test or inspect all units or components being produced. At the user end the same applies, though for electronic components automatic test equipment has fairly recently made 100% testing feasible. An alternative to inspecting every item (while still retaining an effective control on the quality of whatever is being produced or consumed) is to use sampling plans. Such plans are designed to assess the quality of a random sample on the assumption that the sample quality is indicative of the batch quality. (Some prefer the synonymous word "lot" to "batch".) There are many ways to sample; only a couple are discussed here. The various plans are based on slightly different theoretical considerations and physical premises, all of which are valid under varying circumstances. It is simply a question of choosing which one best suits your needs.

Inspection tests may be classified as attribute or variable, depending on what type of component characteristic it measures.

Attribute inspection measures some characteristic which is either good or bad (e.g., during a pull test on a diode, the device breaks or remains intact). Variables inspection measures some characteristic which can vary on a continuous scale (e.g., reverse leakage current of a diode). This chapter only considers in detail the attribute tests. Hence, sampling plans applicable to attribute inspection are termed "sampling by attributes."

Before proceeding to the details, a few commonly used terms are defined:

- o Inspection. The process of measuring, examining or testing the unit of product against some requirement, e.g., measuring resistance values against tolerance specification in incoming inspection department.

- o Defective. A unit which is in some way unacceptable, e.g., resistor out-of-tolerance.

- o Defect. A fault. Note that a defective unit may contain more than one defect and this is why the distinction between defect and defective is made. The distinction is used in inspecting units which contain many components (e.g., a resistor has two nicks in the connection wire, either of which may shorten its life. Hence, there are two defects, the component itself is defective and therefore there is only one defective.)

- o Lot or Batch. A collection of production units from which a sample is to be drawn. Typically the units of a lot will have been made under identical conditions so we can usually safely assume that they will be homogeneous in all respects, except for the defectives.

- o Acceptable Quality Level (AQL). A specified quality (expressed in defects per hundred units or percent defective) which the sampling procedure will accept most of the time.

- o Lot Tolerance Percent Defective (LTPD). Some limiting quality beyond which the sampling plan will reject most of the units.

- o Risk. The probability of making a wrong decision.

- o Producer's Risk ($\alpha$). The probability of a good batch being rejected.

- o Consumer's Risk ($\beta$). The probability of a bad batch being accepted.

Evaluation of Batch Quality by Binomial Sampling

Consider a production batch of 1N914 diodes, and suppose that the batch size is 2472. If the (unknown) number of defective diodes in the batch is d then the proportion defective, p, is given by:

$$p = d/2472 \tag{16}$$

and thus the percent defective is $\left(\frac{d}{2472} \times 100\%\right)$

Thus, the probability of drawing (at random) a defective component from the batch is p. In a sample of n diodes, the probability P(r) of drawing r defectives is given by:

$$P(r) = \binom{n}{r} p^r (1 - p)^{n-r} \tag{17}$$

where

$\binom{n}{r}$ is the binomial coefficient for the number of combinations of "r out of n":

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

In fact (17) is the binomial distribution formula. Suppose, for the purposes of this example, we count the number of defectives to see how well the plans work and find d = 52. Then p = 0.021 (from (16)). Thus, if n is now defined as 5, then since p = 0.021, equation (17) may be solved for all r and we find:

$P(0) = 0.899$
$P(1) = 0.097$
$P(2) = 0.004$
$P(3) = 8.92 \times 10^{-5}$
$P(4) = 9.58 \times 10^{-7}$
$P(5) = 4.11 \times 10^{-9}$

So in a set of 50 random samples of 5 diodes, the following table summarizes the number of defectives which the theory would lead us to expect to be drawn in each sample.

TABLE 12: EXPECTED NUMBER OF DEFECTIVES PER SAMPLE

| Number of Defectives Drawn (r) per Sample | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Expectation (50 x P(r)) | 44.95 | 4.83 | 0.21 | 0.004 | $<10^{-4}$ | $<10^{-6}$ |

Thus, for example, of the 50 samples, we would expect 45 to contain no defectives, 5 to contain one defective, and more than one defective to be less likely. In fact, when the sample were drawn, the following was observed:

TABLE 13: OBSERVED NUMBER OF DEFECTIVES

| Number of Defectives | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Actual Observations | 46 | 4 | 0 | 0 | 0 | 0 |

Though there is some discrepancy, the observations follow the theoretical expectation quite well. To further study the random fluctuations, this little experiment was repeated ten times, with the following results:

TABLE 14: REPEATED SAMPLES

| Number of Defectives | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Observations: | | | | | | |
| 1st set of 50 samples | 46 | 4 | 0 | 0 | 0 | 0 |
| 2nd set of 50 samples | 44 | 6 | 0 | 0 | 0 | 0 |
| 3rd set of 50 samples | 42 | 7 | 1 | 0 | 0 | 0 |
| 4th set of 50 samples | 44 | 6 | 0 | 0 | 0 | 0 |
| 5th set of 50 samples | 46 | 4 | 0 | 0 | 0 | 0 |
| 6th set of 50 samples | 42 | 8 | 0 | 0 | 0 | 0 |
| 7th set of 50 samples | 45 | 5 | 0 | 0 | 0 | 0 |
| 8th set of 50 samples | 44 | 6 | 0 | 0 | 0 | 0 |
| 9th set of 50 samples | 46 | 4 | 0 | 0 | 0 | 0 |
| 10th set of 50 samples | 45 | 5 | 0 | 0 | 0 | 0 |

Having established that defectives within a sample seem to follow the theory quite well, it may be seen that sampling plans could be defined to exploit these probabilistic laws. The example of drawing fifty samples was used to clarify; however, practical sampling plans are often defined on a single sample (usually larger than five). Before introducing such plans, it is necessary to introduce the concept of risk.

## Producer and Consumer Risks

During testing of the 1N914 diodes in the producer's plants, the sampling plan employed will, for the most part, accept good batches and reject bad batches. Unless 100% inspection is employed, any sampling plan, no matter how good, will also have a risk of rejecting some good batches and accepting some bad ones. These risks are defined as producer and consumer risks, $\alpha$ and $\beta$, as follows:

Producer's risk = [ Probability of rejecting a good batch ] = $\alpha$
Consumer's risk = [ Probability of accepting a bad batch ] = $\beta$

The concept can be laid out (if a little naively) as a table of consequences:

TABLE 15: CONSEQUENCES AND RISKS

| Sampling Plan Decision → / Actual ↓ | Accept Batch | Reject Batch |
|---|---|---|
| Batch good | Everyone happy | Producer wastes good product. Risk = $\alpha$ |
| Batch bad | Consumer uses bad product; inconvenience and possible rework expense. Risk = $\beta$ | Everyone happy |

(Recall Chapter 0 and notice the relation to hypothesis testing, significance and Types I and II error. Here the null hypothesis is that the batch is good.)

Note that, in reality, the consequences are not so clearly defined. If bad batches are accepted, the risk is defined as the consumer's, although in reality the consumer is likely to ensure that the producer's day is also ruined.

How Large Should the Sample Be?

Producer and consumer risks vary according to the size of the sample drawn. Though it is clear that the sample size (used in the exercise earlier) of five is inadequate for a single sample on which to base a decision about a large batch of electronic components, it is not clear how large the sample should be. Commonly used industrial sampling plans (like LTPD and AQL, discussed later) determine the sample size according to some risk criteria to minimize the expense of sampling.

LTPD Plans

The worst quality the consumer is prepared to accept is defined as the Lot Tolerance Percent Defective (LTPD). LTPD sampling plans have been designed which try hard to reject batches having quality worse than the LTPD. Their theoretical derivation is along similar lines (though less crude) to the simple binomial example given earlier. The plans set the consumer risk nominally at 10%. A set of LTPD sampling plans is reproduced (from Ref. 16) in Table 16, which gives the minimum sample size required to assess a given LTPD requirement. The various LTPD options are set out along the top. The options are 50%, 30%, 20%, 15%, 10%, 7% and so on through three decades of the sequence, which was chosen for convenience. The acceptance number C is given down the left-hand side of the table: we accept our batch if the sample has no more than C defectives (and reject if it has more than C). Note that Table 16 is simplified from the original work (Ref. 34).

| Max. Percent Defective (LTPD) or λ | 50 | 30 | 20 | 15 | 10 | 7 | 5 | 3 | 2 | 1.5 | 1 | 0.7 | 0.5 | 0.3 | 0.2 | 0.15 | 0.1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Acceptance Number (c) (r = c + 1) | Minimum Sample Sizes (For device-hours required for life test, multiply by 1000) | | | | | | | | | | | | | | | | |
| 0 | 5 (1.03) | 8 (0.64) | 11 (0.46) | 15 (0.34) | 22 (0.23) | 32 (0.16) | 45 (0.11) | 76 (0.07) | 116 (0.04) | 153 (0.03) | 231 (0.02) | 328 (0.02) | 461 (0.01) | 767 (0.007) | 1152 (0.005) | 1534 (0.003) | 2303 (0.002) |
| 1 | 8 (4.4) | 13 (2.7) | 18 (2.0) | 25 (1.4) | 38 (0.94) | 55 (0.65) | 77 (0.46) | 129 (0.28) | 195 (0.18) | 258 (0.14) | 390 (0.09) | 555 (0.06) | 778 (0.045) | 1296 (0.027) | 1946 (0.018) | 2592 (0.013) | 3891 (0.009) |
| 2 | 11 (7.4) | 18 (4.5) | 25 (3.4) | 34 (2.24) | 52 (1.6) | 75 (1.1) | 105 (0.78) | 176 (0.47) | 266 (0.31) | 354 (0.23) | 533 (0.15) | 759 (0.11) | 1065 (0.080) | 1773 (0.045) | 2662 (0.031) | 3547 (0.022) | 5323 (0.015) |
| 3 | 13 (10.5) | 22 (6.2) | 32 (4.4) | 43 (3.2) | 65 (2.1) | 94 (1.5) | 132 (1.0) | 221 (0.62) | 333 (0.41) | 444 (0.31) | 668 (0.20) | 953 (0.14) | 1337 (0.10) | 2226 (0.062) | 3341 (0.041) | 4452 (0.031) | 6681 (0.018) |
| 4 | 16 (12.3) | 27 (7.3) | 38 (5.3) | 52 (3.9) | 78 (2.6) | 113 (1.8) | 158 (1.3) | 265 (0.75) | 398 (0.50) | 531 (0.37) | 798 (0.25) | 1140 (0.17) | 1599 (0.12) | 2663 (0.074) | 3997 (0.049) | 5327 (0.037) | 7994 (0.025) |
| 5 | 19 (13.8) | 31 (8.4) | 45 (6.0) | 60 (4.4) | 91 (2.9) | 131 (2.0) | 184 (1.4) | 308 (0.85) | 462 (0.57) | 617 (0.42) | 927 (0.28) | 1323 (0.20) | 1855 (0.14) | 3090 (0.085) | 4638 (0.056) | 6181 (0.042) | 9275 (0.028) |
| 6 | 21 (15.6) | 35 (9.4) | 51 (6.6) | 68 (4.9) | 104 (3.2) | 149 (2.2) | 209 (1.6) | 349 (0.94) | 528 (0.62) | 700 (0.47) | 1054 (0.31) | 1503 (0.22) | 2107 (0.155) | 3509 (0.093) | 5267 (0.062) | 7019 (0.047) | 10533 (0.031) |
| 7 | 24 (16.6) | 39 (10.2) | 57 (7.2) | 77 (5.3) | 116 (3.5) | 166 (2.4) | 234 (1.7) | 390 (1.0) | 589 (0.67) | 783 (0.51) | 1178 (0.34) | 1680 (0.24) | 2355 (0.17) | 3922 (0.101) | 5886 (0.067) | 7845 (0.051) | 11771 (0.034) |
| 8 | 26 (18.1) | 43 (10.9) | 63 (7.7) | 85 (5.6) | 126 (3.7) | 184 (2.6) | 258 (1.8) | 431 (1.1) | 648 (0.72) | 864 (0.54) | 1300 (0.36) | 1854 (0.25) | 2599 (0.18) | 4329 (0.108) | 6498 (0.072) | 8660 (0.054) | 12995 (0.036) |
| 9 | 28 (19.4) | 47 (11.5) | 69 (8.1) | 93 (6.0) | 140 (3.9) | 201 (2.7) | 282 (1.9) | 471 (1.2) | 709 (0.77) | 945 (0.58) | 1421 (0.38) | 2027 (0.27) | 2842 (0.19) | 4733 (0.114) | 7103 (0.077) | 9468 (0.057) | 14206 (0.038) |
| 10 | 31 (19.9) | 51 (12.1) | 75 (8.4) | 100 (6.3) | 152 (4.1) | 218 (2.9) | 306 (2.0) | 511 (1.2) | 770 (0.80) | 1025 (0.60) | 1541 (0.40) | 2199 (0.28) | 3082 (0.20) | 5133 (0.120) | 7704 (0.080) | 10268 (0.060) | 15407 (0.040) |
| 11 | 33 (21.0) | 54 (12.8) | 83 (8.3) | 111 (6.2) | 166 (4.2) | 236 (2.9) | 332 (2.1) | 555 (1.2) | 832 (0.83) | 1109 (0.62) | 1664 (0.42) | 2378 (0.29) | 3323 (0.21) | 5546 (0.12) | 8319 (0.083) | 11092 (0.062) | 16638 (0.042) |
| 12 | 36 (21.4) | 59 (13.0) | 89 (8.6) | 119 (6.5) | 178 (4.3) | 254 (3.0) | 356 (2.2) | 594 (1.3) | 890 (0.86) | 1187 (0.65) | 1781 (0.43) | 2544 (0.3) | 3562 (0.22) | 5936 (0.13) | 8904 (0.086) | 11872 (0.065) | 17808 (0.043) |
| 13 | 38 (22.3) | 63 (13.4) | 95 (8.9) | 126 (6.7) | 190 (4.5) | 271 (3.1) | 379 (2.26) | 632 (1.3) | 948 (0.89) | 1264 (0.67) | 1896 (0.44) | 2709 (0.31) | 3793 (0.22) | 6321 (0.134) | 9482 (0.089) | 12643 (0.067) | 18964 (0.045) |
| 14 | 40 (23.1) | 67 (13.8) | 101 (9.2) | 134 (6.9) | 201 (4.6) | 286 (3.2) | 403 (2.3) | 672 (1.4) | 1007 (0.92) | 1343 (0.69) | 2015 (0.46) | 2878 (0.32) | 4029 (0.23) | 6716 (0.138) | 10073 (0.092) | 13431 (0.069) | 20146 (0.046) |
| 15 | 43 (23.3) | 71 (14.1) | 107 (9.4) | 142 (7.1) | 213 (4.7) | 305 (3.3) | 426 (2.36) | 711 (1.41) | 1066 (0.94) | 1422 (0.71) | 2133 (0.47) | 3046 (0.33) | 4265 (0.235) | 7108 (0.141) | 10662 (0.094) | 14216 (0.070) | 21324 (0.047) |
| 16 | 45 (24.1) | 74 (14.9) | 112 (9.7) | 150 (7.2) | 225 (4.8) | 321 (3.37) | 450 (2.41) | 750 (1.44) | 1124 (0.96) | 1499 (0.72) | 2249 (0.48) | 3212 (0.337) | 4497 (0.241) | 7496 (0.144) | 11244 (0.096) | 14992 (0.072) | 22487 (0.048) |
| 17 | 47 (24.7) | 79 (14.7) | 118 (9.86) | 158 (7.36) | 236 (4.93) | 338 (3.44) | 473 (2.46) | 788 (1.48) | 1182 (0.98) | 1576 (0.74) | 2364 (0.49) | 3377 (0.344) | 4728 (0.246) | 7880 (0.148) | 11812 (0.098) | 15759 (0.074) | 23639 (0.049) |
| 18 | 50 (24.9) | 83 (15.0) | 124 (10.0) | 165 (7.54) | 248 (5.02) | 354 (3.51) | 496 (2.51) | 826 (1.51) | 1239 (1.0) | 1652 (0.75) | 2478 (0.50) | 3540 (0.351) | 4956 (0.251) | 8260 (0.151) | 12390 (0.100) | 16520 (0.075) | 24780 (0.050) |
| 19 | 52 (25.5) | 86 (15.4) | 130 (10.2) | 173 (7.76) | 259 (5.12) | 370 (3.58) | 518 (2.56) | 864 (1.53) | 1296 (1.02) | 1728 (0.77) | 2591 (0.52) | 3702 (0.358) | 5183 (0.256) | 8638 (0.153) | 12957 (0.102) | 17276 (0.077) | 25914 (0.051) |
| 20 | 54 (26.1) | 90 (15.6) | 135 (10.4) | 180 (7.82) | 271 (5.19) | 386 (3.65) | 541 (2.60) | 902 (1.56) | 1353 (1.04) | 1803 (0.78) | 2705 (0.52) | 3864 (0.364) | 5410 (0.260) | 9017 (0.156) | 13526 (0.104) | 18034 (0.078) | 27051 (0.052) |
| 25 | 65 (27.0) | 109 (16.1) | 163 (10.8) | 217 (8.08) | 326 (5.38) | 466 (3.76) | 652 (2.69) | 1086 (1.61) | 1629 (1.08) | 2173 (0.807) | 3259 (0.538) | 4656 (0.376) | 6518 (0.269) | 10863 (0.161) | 16295 (0.108) | 21726 (0.081) | 32589 (0.054) |

TABLE 16: SAMPLING TABLE (Reproduced from MIL-M-38510E)

## Example

Recall the batch of 1N914 diodes and that their actual percent defective is 2.1%. If we define an LTPD of 2(%), then the worst acceptable quality is 2%. Thus, the LTPD plans should reject the batch (as having a quality worse than 2%). Since we accept a consumer risk we expect some samples to (incorrectly) accept the batch (nominally 1 in 10). Entering the table at an LTPD of 2, the sample size is defined as 116 with an accept/reject criteria of 0/1. In ten such samples, the following results were yielded:

| Sample No. | Number of Defective Drawn | Decision |
|---|---|---|
| 1 | 3 | Reject |
| 2 | 2 | Reject |
| 3 | 3 | Reject |
| 4 | 4 | Reject |
| 5 | 2 | Reject |
| 6 | 6 | Reject |
| 7 | 2 | Reject |
| 8 | 3 | Reject |
| 9 | 2 | Reject |
| 10 | 3 | Reject |

## Small Batch Procedures

Notice that the LTPD table defines sample sizes independently of batch size. Intuitively it would seem that the batch size should matter, and in fact it does. However, the effect is negligible for "large" batch sizes, and it is found that the batch size really starts to have an effect at batch sizes of around 200. For this reason, hypergeometric plans may be necessary; such plans allow for the effect (on % defective) of removing a sample from a small batch. The LTPDs

given for hypergeometric plans are outside the standard convenient range (because otherwise, the plan would have us pulling weird sample sizes like 7.5 or 15.3). MIL-S-19500 tackles the problem in a different way by explicitly defining (for small batches) sample size and reject criteria for each test. Because of the differences in component yields and defect rates, MIL-S-19500 defines small batches as less than 500.

## AQL Plans

The acceptable quality level (AQL) defines the minimum quality acceptable (as the process average). Thus, AQL plans try hard to accept batches having quality equal to or better than the AQL. AQL schemes are defined in MIL-STD-105 and are set up to have a producer's risk of nominally 5%. Though the sample sizes given by the plans of MIL-STD-105 do depend on the batch size, their use is similar to LTPD, with a sample size and an accept/reject criteria being defined for required AQL.

## Operating Characteristic Curves

The effectiveness of a sampling plan may be assessed from its operating characteristic (O-C) curve. Consider an ideal sampling plan; then some quality $Q_0$ could be defined such that all batches with quality of at least $Q_0$ are accepted and those with quality worse then $Q_0$ are rejected.

Then the probability of accepting a batch of given quality Q (in terms of % defective) is defined by Figure 14, the ideal O-C curve. There are no producer or consumer risks.

In reality, $Q_0$ is generally not a sharply defined point but a grey area between two values $Q_1$ and $Q_2$. Furthermore, batches of at least (the higher) quality $Q_1$ have a (producer) risk of rejection and batches of quality $Q_2$ or worse have risk (to the consumer) of acceptance, as Figure 15 illustrates.

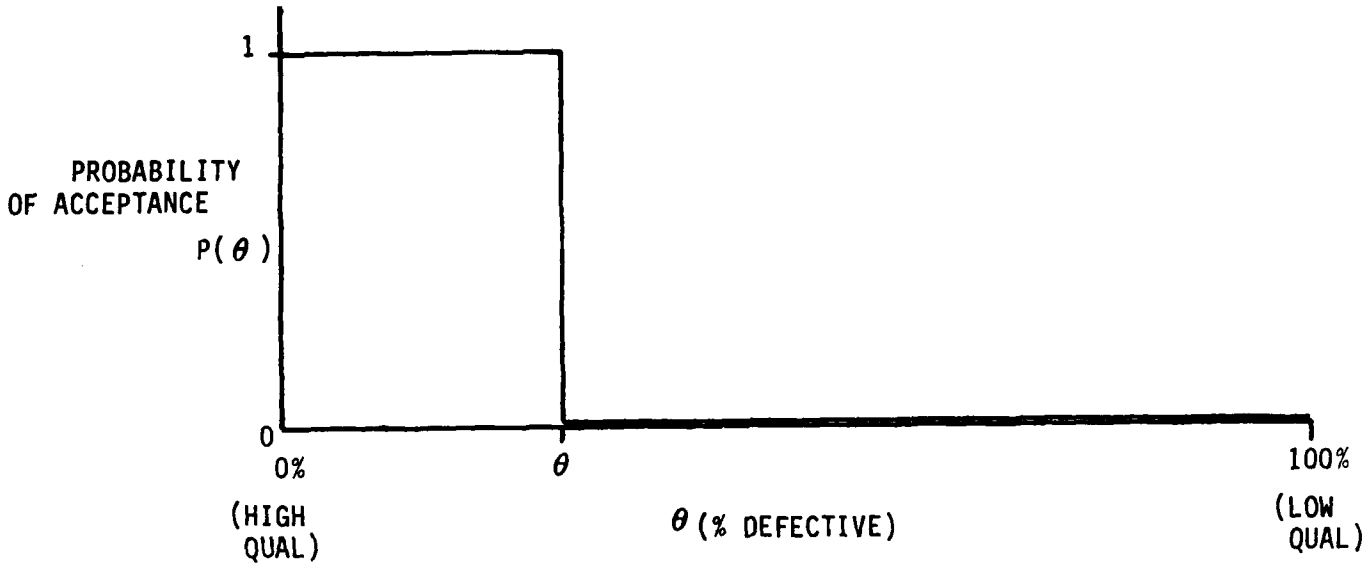



FIGURE 14: IDEAL O-C CURVE

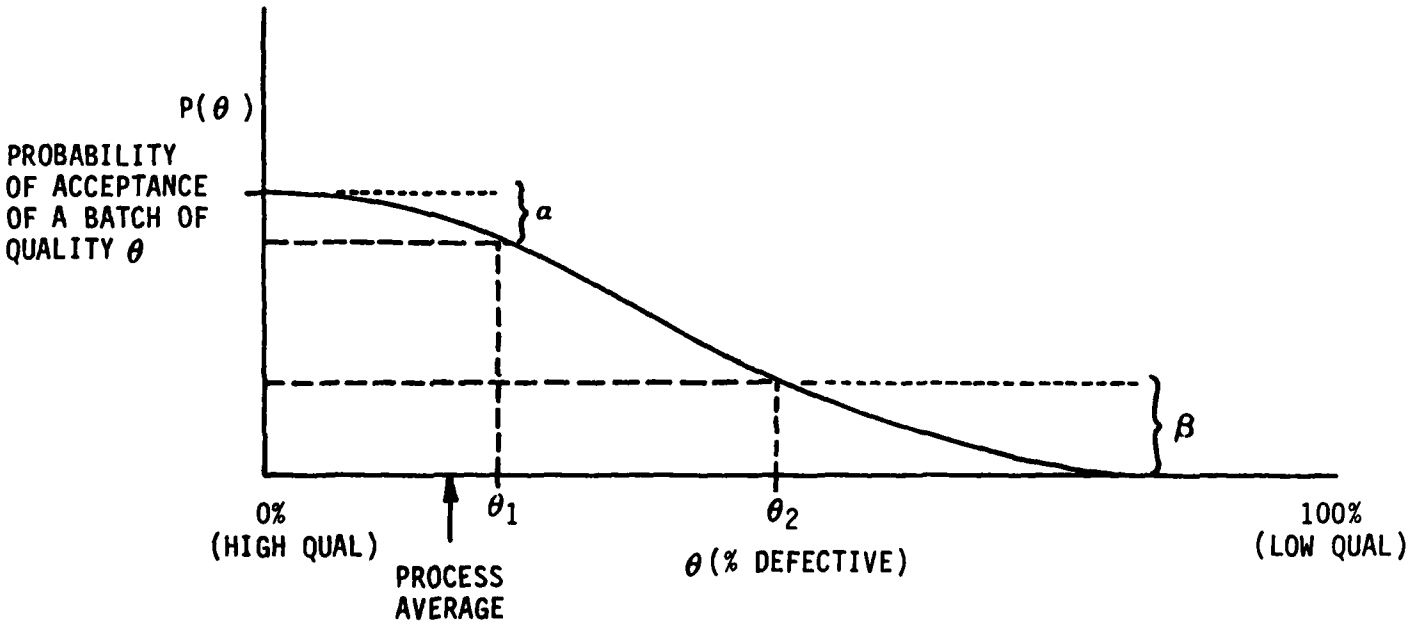



FIGURE 15: PRACTICAL O-C CURVE

For LTPD schemes, the LTPD is defined at $Q_2$. For AQL schemes, the AQL is defined at $Q_1$. The process average is better than both, as indicated.

## Single, Double and Multiple Sampling Plans

We have already seen that producer and consumer risks exist at levels $\alpha$ and $\beta$ for single-sample plans. Double-sampling plans can be used to reduce these risks by making use of compound probabilities. MIL-M-38510 and MIL-S-19500 include clauses to the effect that a second sample may be taken if the first is rejected on the grounds that it may have caught the (producer's) risk. The second sample is pooled with the first, and the decision whether to accept or reject is made on the whole sample using the accept/reject criteria for that sample size. At first glance, some have expressed concern that this is simply trying again until the answer that was really wanted is found. In fact, this is not so: the double sample merely establishes greater confidence (reduces risk of error) that the truth is reflected in the sample. An example further clarifies. Suppose (with the 1N914 diodes again) that an LTPD of 5% is specified. Then the sample size is (initially) defined from Table 16 as 45. If the sample of 45 contains a defect, the batch is rejected by the plan since the acceptance number (C) is zero. However, if the producer wishes, a second sample of 32 may be drawn to make the total sample up to 77 (the next line down for an LTPD of 5). The acceptance number for the whole sample is now 1, and thus if there are no more defects found the batch is accepted. If the batch had truly been "bad," a second defect would have been likely, to confirm the earlier result.

Multiple plans also exist which further compound repeated samples. The penalty for the higher confidence is, of course, greater expense.

## Comparison of LTPD and AQL Schemes

Both LTPD and AQL plans perform essentially the same function; they just approach it differently. The LTPD plan given in Table 16 includes a number in parentheses below each sample size. This gives an approximate equivalence between the two plans, though it is difficult to make direct comparisons. In other words, the number in parentheses is an approximate AQL which would be obtained under the sample conditions defined for the LTPD plan.

Table 17 summarizes the major differences between LTPD and AQL schemes.

TABLE 17: COMPARISON OF AQL AND LTPD

| Aspect | LTPD | AQL |
|---|---|---|
| Risk | Consumer risk set at 10% | Producer risk set at 5% |
| Sample Size | Independent of batch size (above 200) | Dependent on batch size |
| Concept | Tries hard to reject bad batches | Tries hard to accept good batches |
| Where called up | MIL-M-38510<br>MIL-S-19500 | MIL-STD-105 |
| Applications | Microcircuits and solid state | Passives |

## Applications in the Electronic Component Industry

Notice that in the 1N914 diode example used throughout this text it was assumed that a single inspection test was chosen to evaluate the batch. In practice, it is usual for a number of tests to be applied to

evaluate a batch. The microcircuit test method specifications of MIL-STD-883 invoke a range of inspection tests in four groups (A, B, C and D), each of which tests a different aspect of the device (electrical, environmental, package and die-related). An inspection lot is thus dividable into sublots for undergoing each of the tests called up for that device, and this can be effected simultaneously. For a batch to be accepted, it is required to be accepted by all of the tests (though they will not generally all be performed on the same sublot). MIL-STD-750 (for semiconductors) is similar. Sometimes, LTPD may be defined for a group of tests together.

In some cases, these specifications override the LTPD plans (e.g., where destructive testing would be too expensive for the sample defined). Then the sample will be defined in the device "slash" sheet. For example, an entry of the following form

2(0)

means the sample size is two, the accept number is zero.

LTPD plans may be extended to accomodate life testing. Though life testing is a variables test (time is continuous) it can be made an attribute test by specifying a life test period of 1000 hours and accepting (or rejecting) components which have survived (or failed before) 1000 hours. Table 16 may be used to define the total number of part hours required for a life test to establish confidence in a component failure rate by multiplying sample size by 1000 hours.

# CHAPTER 6

## Weibull Plotting and Analysis

CHAPTER 6

# Weibull Plotting and Analysis

## Summary

This note gathers together some useful techniques associated with Weibull plots and the associated data analysis. Methods for plotting from scratch are first given and the ideas developed so that the reader may perform failure data analysis. Goodness-of-fit tests are recommended and examples are given.

## Introduction

The Weibull distribution is particularly useful in analyzing life data since it takes on the form (or approximate form, in some cases) of many other statistical distributions used to model life.

The form of the Weibull varies between texts but a common one is given by the probability density function (p.d.f.):

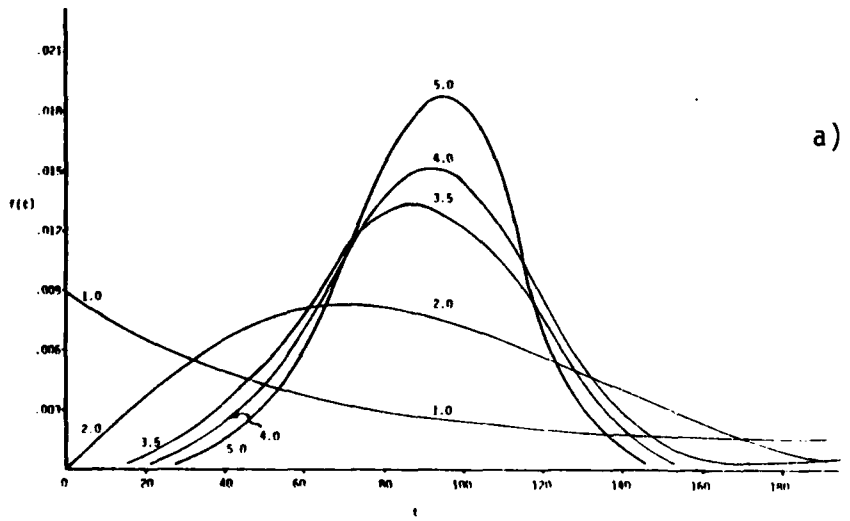
$$f(t) = \frac{\beta}{\alpha}\left(\frac{t-\gamma}{\alpha}\right)^{\beta-1} \exp -\left(\frac{t-\gamma}{\alpha}\right)^{\beta} \qquad (18)$$

where

$\alpha$ is a scale parameter (characteristic life)
$\beta$ is a shape parameter
$\gamma$ is the minimum life

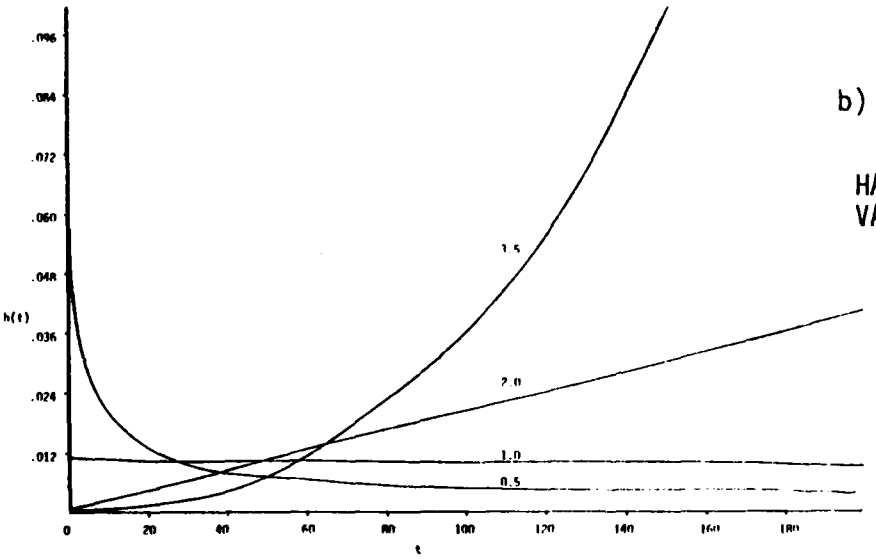
In practice, $\gamma$ is often taken as or found to be zero, and the p.d.f. becomes simpler.

The Weibull p.d.f. is sketched for five values of $\beta$. A full range of sketches is given in Figure 16(a).



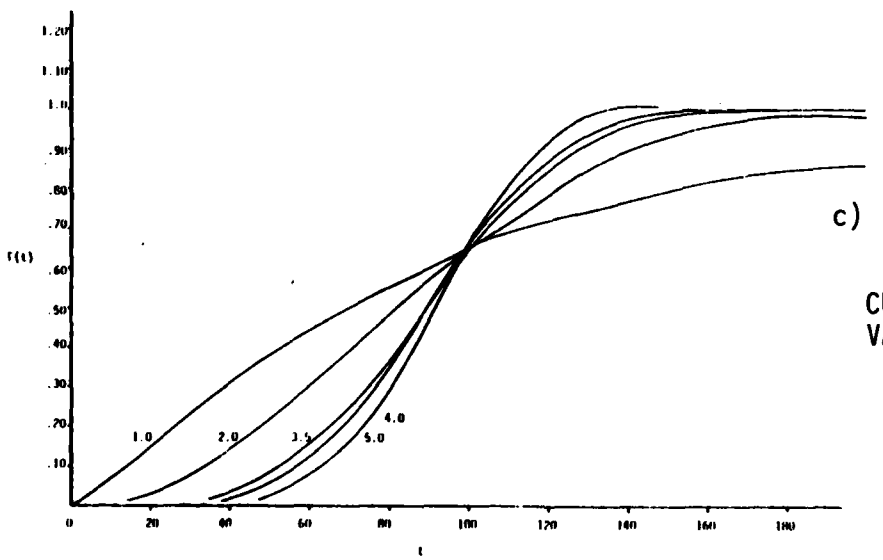

a)

DENSITY, WEIBULL,
VARIOUS SHAPES ( $\beta$ )




b)

HAZARD, WEIBULL,
VARIOUS SHAPES ( $\beta$ )




c)

CUMULATIVE, WEIBULL,
VARIOUS SHAPES ( $\beta$ )

FIGURE 16: WEIBULL DENSITIES, HAZARDS AND CUMULATIVE DENSITIES

Depending on the values of $\beta$, the Weibull takes on the shape of other distributions as follows:

TABLE 18: WEIBULL SHAPE PARAMETERS

| Beta | Distribution Type |
|---|---|
| $\beta < 1$ | Gamma ($k < 1$) |
| $\beta = 1$ | Exponential |
| $\beta = 2$ | Rayleigh |
| $\beta = 3.44$ | Normal (approx.) |

In addition, with a transformation the Weibull distribution also describes the family of extreme value distributions. The concept of the hazard function is also useful at this stage. The hazard function is defined mathematically as

$$h(t) = \frac{f(t)}{1 - F(t)} \tag{19}$$

where $f(t)$ is the failure distribution and $F(t)$ is the associated cumulative density function. Thus, $h(t)$ is an expression of the number of parts failing at time t, as a function of the surviving parts. For the Weibull distribution, $h(t)$ is a decreasing function for $\beta < 1$, increasing for $\beta > 1$, and constant for $\beta = 1$. Further details are given in Figure 16(b).

There is a vast treatment of Weibull theory in the literature but this brief summary will suffice to understand and construct probability plots.

Weibull Plotting

Probability paper is similar to logarithmic and other scale-transforming graph paper, except that it transforms the probability

scale. The other axis is defined by the random variable. The transformed scale is such that a set of data from a given distribution will appear as a straight line when plotted as coordinate points.

Probability paper is available for the Weibull distribution. Because of the attractive property noted above that the Weibull takes on many other distribution shapes, all we need to do is plot the data on Weibull probability paper, and if a straight line results there is some indication that the data is Weibull. Simple estimation of $\alpha$ and $\beta$ are given on the paper so that we can further identify which Weibull our data follows (exponential, Rayleigh or whatever). In practice, the process is not quite as clean since data rarely fall exactly on a straight line. Statistical tests are advisable to minimize the risk of erroneous conclusions.

A typical piece of Weibull paper is shown in Figure 17.

A detailed method for time-to-failure (TTF) data is now given:

(i) Order the TTF data smallest to largest and assign a rank to each observation. So if there are n observations, the ranks will run from 1 to n. We will call "i" the rank of the $i^{th}$ observation. If any observations are found to be equal, then the median rank is used for each of those "tied" observations.

(ii) The observed cumulative distribution is then evaluated by calculating the percentiles for each observation. There are a number of ways to do this, the simplest being:

$$P_i = \frac{i}{n} \times 100\%$$

where $P_i$ is the cumulative percent of failures at the $i^{th}$ observation.

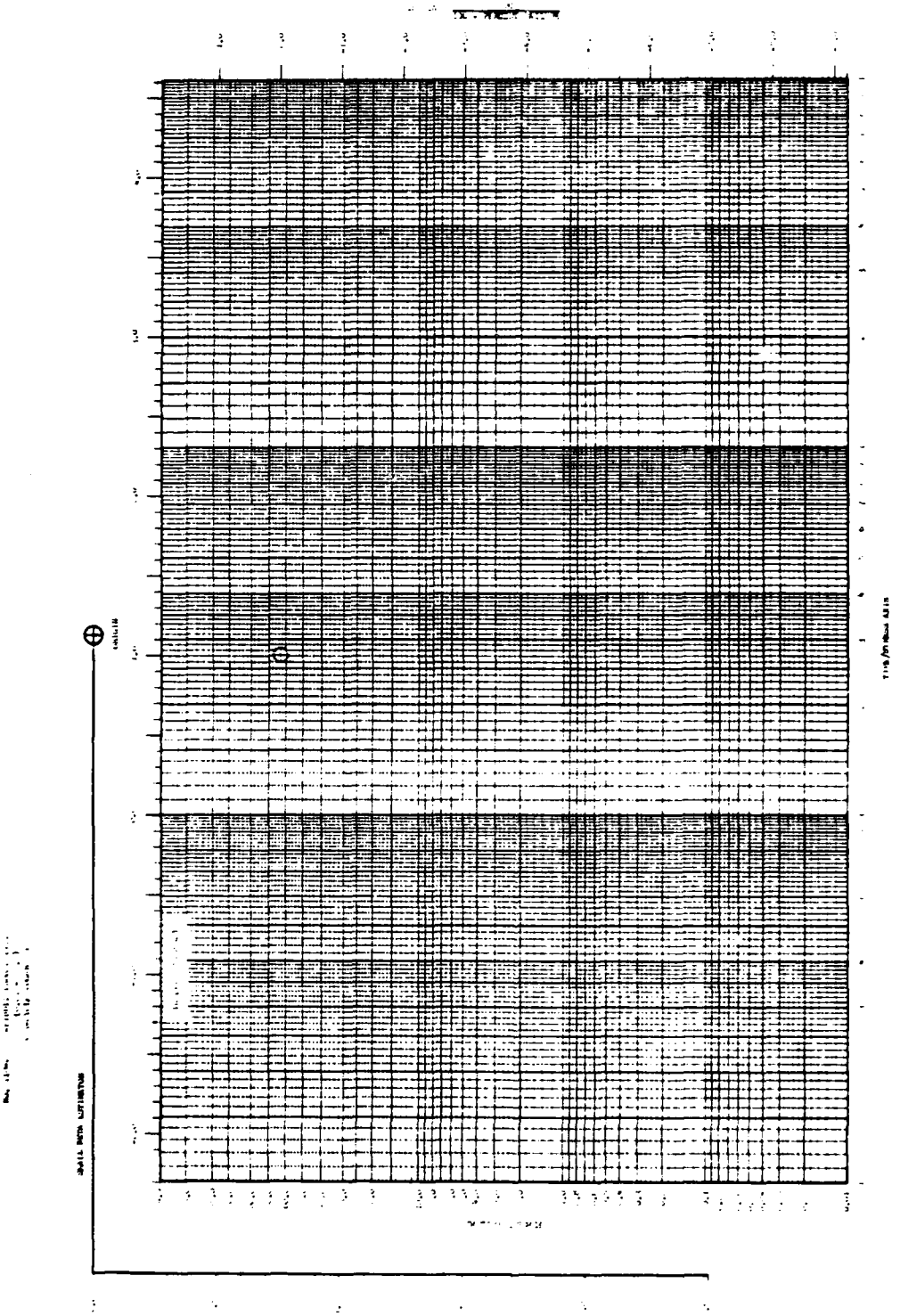


FIGURE 17: WEIBULL PAPER

Unfortunately, this method is very biased. The bias may be intuitively appreciated by taking the $n^{th}$ observation, finding $P_i = 100\%$ and noting that we are unlikely to have 100% of a theoretical distribution in a single sample. $100\ (i/(n + 1))$ is better but is still unduly pessimistic. A simple unbiased method is given by Bernard's formula:

$$P_i = \left(\frac{i - 0.3}{n + 0.4}\right)$$

This method is used from hereon.

(iii) Plot the values of $P_i$ against the TTF on Weibull probability paper.

(iv) If a reasonably straight line results, put in the line of best fit by eye. (Optimization methods may be used if necessary; see Chapter 6, Optimization Methods). If the data obviously are not in a straight line then the data is probably not Weibull (or $\gamma \neq 0$, see Chapter 6, What if a Straight Line Does Not Result?).

(v) Read off the value of $\beta$ on the scale provided. The methods for doing this vary depending on whose paper is used. Most construct either a parallel or perpendicular on the fitted line, through some origin provided. Each method will be self-explained by the paper.

$\alpha$ (the characteristic life) may be estimated from the value of TTF corresponding to P = 63.2%, as determined by the fitted line, i.e., construct a horizontal through the 63.2 percentile and drop a perpendicular onto the TTF axis, from the point where the horizontal intersects the fitted line. Most Weibull paper includes a clear indication of the 63.2% line.

Thus the characteristic life of a component is the time at which 63.2% of the population have failed. (It is analogous to the characteristic time of a capacitor).

The mean life or mean-time-to-failure (MTTF) is sometimes accomodated on Weibull paper. It may be read from a percentile (as for $\alpha$) but the percentile varies according to the value of $\beta$.

An approximate list of percentiles which may be used to evaluate the MTTF is given in Table 19.

TABLE 19: RELATION BETWEEN $\beta$ AND MTTF

| $\beta$ | Percentile Used to Estimate Mean Life |
|---|---|
| 0.5 | 75 % |
| 1.0 | 63.2% |
| 1.5 | 57.5% |
| 2.0 | 54.5% |
| 2.5 | 52.5% |
| 3.0 | 51 % |
| 3.44 | 50 % |

The analytic relation is given by:

$$\text{MTTR} = \{\Gamma(1 + 1/\beta\}\alpha \qquad (20)$$

Example

An example of the use of Weibull paper is now given. Eighteen observations of TTF were collected and are ordered in Table 20. The percentiles, $P_i$, are estimated using Bernard's formula.

The data is now plotted on Weibull probability paper in Figure 18.

The line is fitted as shown and found to have a $\beta$ of about 1.16 as shown by the (dotted) parallel line. A value of 1.16 indicates that the

TABLE 20: DATA, RANKS AND PERCENTILES

| TTF ($\times 10^4$ hours) | Rank (i) | $P_i$(%) |
|---|---|---|
| 0.02 | 1 | 3.8 |
| 0.45 | 2 | 9.2 |
| 0.52 | 3 | 14.7 |
| 0.60 | 4 | 20.1 |
| 0.99 | 5 | 25.5 |
| 1.35 | 6 | 31.0 |
| 1.78 | 7 | 36.4 |
| 1.83 | 8 | 41.8 |
| 2.16 | 9 | 47.3 |
| 2.22 | 10 | 52.7 |
| 2.34 | 11 | 58.2 |
| 2.94 | 12 | 63.6 |
| 3.43 | 13 | 69.0 |
| 4.82 | 14 | 74.5 |
| 5.32 | 15½ | 82.6 |
| 5.32 | 15½ | 82.6 |
| 6.44 | 17 | 90.8 |
| 7.8 | 18 | 96.2 |

data are close to exponential. One point appears to be a freak as shown. This is checked later.

From the 63.2% estimated, $\alpha$ is found to be $3.25 \times 10^4$. Hence, the characteristic life is estimated at 32,500 hours.

The mean life (or MTTF) is estimated from (approximately) the 61% value of P and is hence found to be 31,000 hours. Note that if the data were truly exponential, the mean and characteristic life $\alpha$ would be one

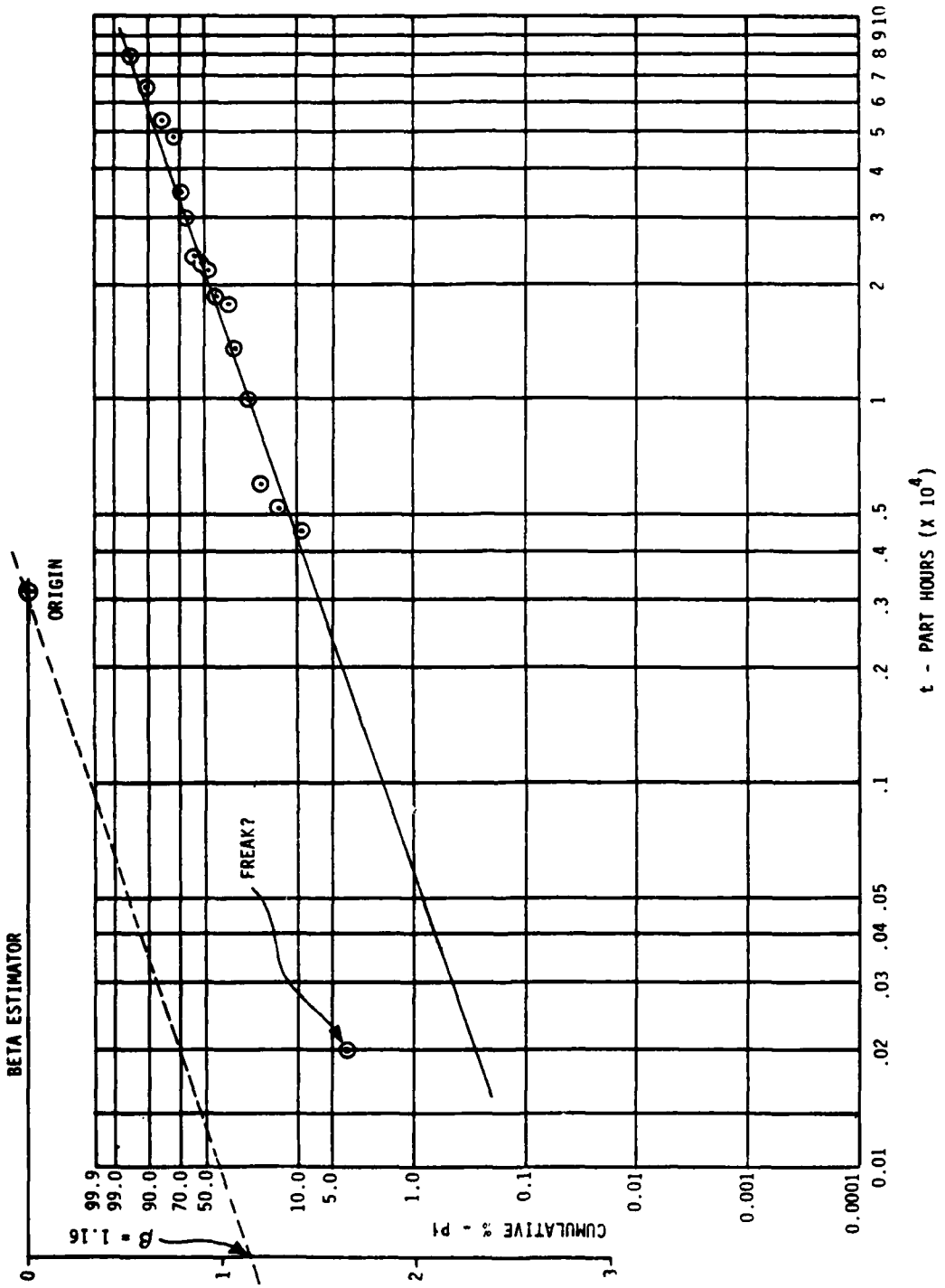



FIGURE 18: WEIBULL PLOT

and the same. Analytically this follows, since from (20)

$$MTTF/\alpha = \Gamma(1 + 1/1) = \Gamma 2 = 1$$

Although the various parameters may be estimated analytically directly from the data, the Weibull probability plot provides a quick, easy method which may be particularly attractive to the non-specialist.

It is emphasised that conclusions based on Weibull plots should strictly be backed up by a statistical goodness-of-fit test such as Kolmogorov - Smirnov or chi-squared tests. The Kolmogorov - Smirnov (K-S) is particularly easy and full details are given in Chapter 4 (Goodness of Fit Tests). In the example, Figure 18 shows the K-S statistic to be about 10% (i.e., 0.1) at $t = 4.8 \times 10^4$ part hours. Reference to tables shows that the critical value at the 20% level is 0.244; hence, our value is not significant. Hence, the distribution of failures is not significantly different from Weibull with $\beta \simeq 1.16$. (Note that the test is only approximate since it is subject to drawing and reading errors.) It is also now possible to evaluate the suspected freak. With a deviation from the straight line of about 3.6% (i.e., 0.036) this point is far less extreme than many of the points which appear to fall very close to the line. This highlights the deceptiveness of data when not judged by a statistic.

What if a straight line does not result?

If the plot describes a curve it may be that the data is still Weibull but that $\alpha$ is not zero. If the line is curved, then a period of dormancy before use or alternatively a period of burn-in (or similar) may be indicated depending on whether the line curves downward or upward. $\alpha$ would then represent the dormancy period, or the burn-in effect. In general, a non-zero gamma may indicate operation in a mixed environment. Methods for estimating $\gamma$ are given in Ref. 8, or one may

estimate by eye from the plot. Note that a curved line does not automatically imply that $\gamma$ is non-zero. It may simply be that the data is not Weibull. Certainly a curve with more than one pronounced inflection (e.g., S-shaped) in it is probably not Weibull.

Necessary caution is as always emphasized and illustrated simply by Figure 19, which is a series of random numbers on Weibull paper. But they are certainly not Weibull (in fact, random numbers are from a uniform distribution which is a special case of the Beta distribution).

Optimization Methods

If it is required to fit the line of best fit using some algorithm rather than by eye to avoid bias; two methods are available:

- o Least squares
- o Minimax

Least squares has the drawbacks that error will not be distributed with equal variance along the line and also that it takes a little time to compute the equations. If a least squares approximation is required, Chapter 8 could be used, though the uneven percentage scale makes this counter-intuitive.

A better solution is to use minimax by fitting the line which gives the minimum value to the maximum % deviation from the line. This could be done approximately by eye, formally by minimax solution (computer programs or numerical methods exist, Ref. 19), or to a good approximation by using the procedure suggested by $M_{age}$ (Ref 20) recently.

If it is imperative that parameter estimates be extremely good, then analytical methods should be used instead of graphical. Such methods are covered in Reference 8.

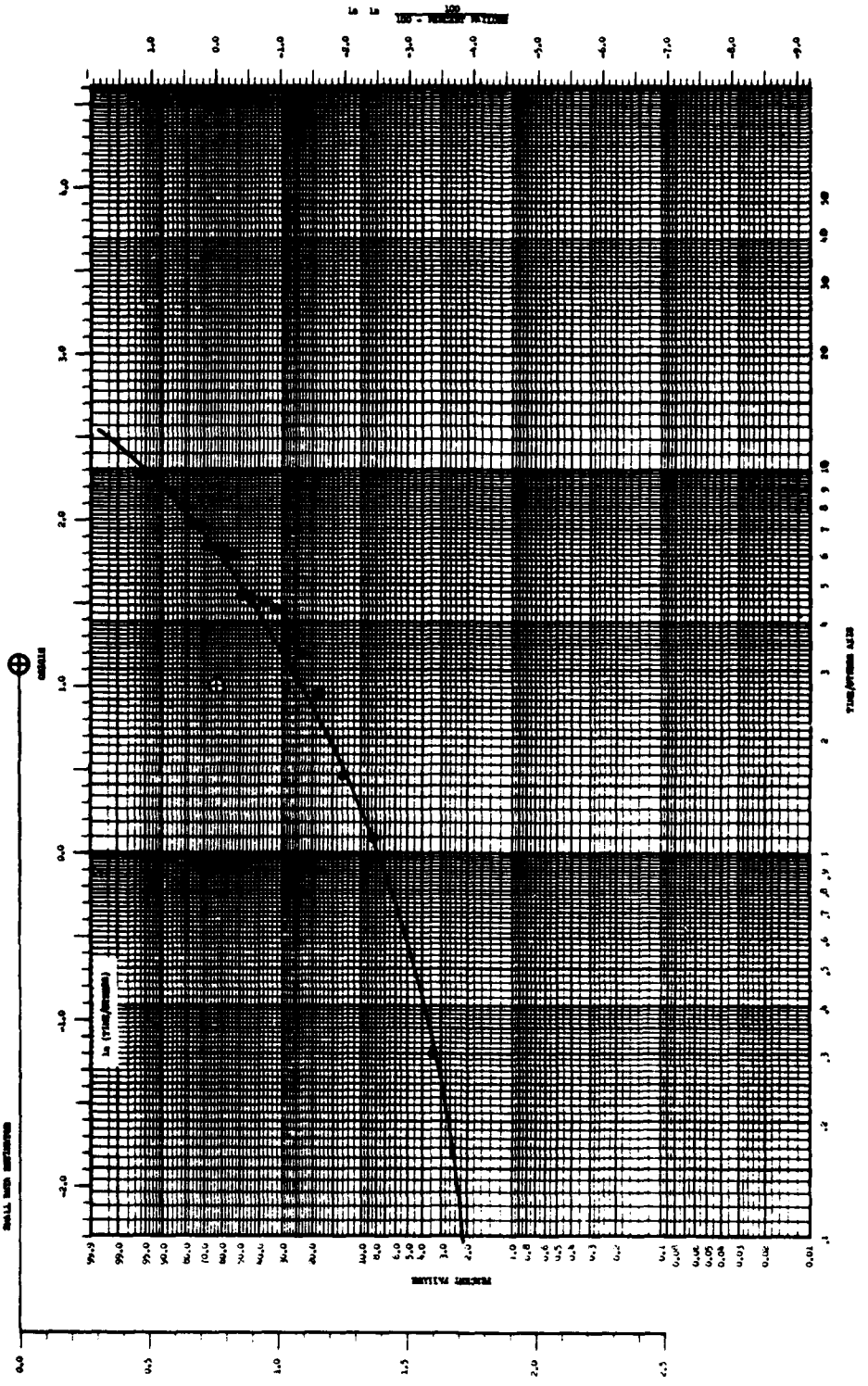


FIGURE 19: RANDOM DIGITS PLOTTED ON WEIBULL PAPER

## Physical Considerations

Information on the failure physics of a component is invaluable in assisting with identifying distributions. In the Weibull distribution, $\beta < 1$ implies a decreasing hazard rate model, $\beta = 1$ implies the constant failure rate, and $\beta > 1$ implies increasing failure rate. For example, vacuum tubes might be expected to last for quite a while and then fail towards the end of their specified life. Then a $\beta > 1$ model would fit the bill.

## Further Methods

It is often necessary to include zero-failure (or survival) data in a Weibull plot. It is also sometimes desirable to construct confidence intervals on $\alpha$, $\beta$, $\gamma$ and various other parameters such as the mean life. L.G. Johnson has written an excellent book (Ref. 13) giving clear non-mathematical techniques for doing all these things and a few more besides.

## Review

Weibull plots are invaluable in analyzing life data; however, care should be taken to back up each plot with a goodness-of-fit test. Small amounts of data ($< 10$?) should be handled with extreme caution.

One problem with this type of work has been discriminating between similar shaped density functions. For example, the exponential distribution can appear virtually identical to the lognormal (with a certain $\sigma$ parameter). The underlying physical principles can help here; additionally, a thorough investigation of all the moments of distribution rather than just its shape may help. In certain cases, a Bayesian test (choosing which of the two candidate distributions is more likely, given the data) may be feasible.

# CHAPTER 7

## Bayesian Statistics

CHAPTER 7

# Bayesian Statistics

## Summary

This chapter illustrates some perliminaries to the vast subject of Bayesian statistics. The areas which have been controversial are highlighted and those parts which are sometimes misunderstood are clarified. References are provided for the reader who wishes to tackle Bayesian methods in any depth. A numerical example is given. The chapter attempts to present the controversial ideas without bias or opinion.

## Introduction

Bayes' theorem may be derived from probabilistic axioms as

$$P[E_k|A] = \frac{P[A|E_k]\ P[E_k]}{\sum_{j=1}^{n} P[A|E_j]\ P[E_j]} \qquad (21)$$

where

$P[E_k|A]$ is the conditional probability of the event $E_k$ happening, given that A has happened

$j = 1,2...n$. $k$ is a particular $j$

$P[A|E_k]$ is the conditional probability of A given $E_k$

$P[E_k]$ is the probability of $E_k$ happen ıg

The $E_k$ are mutually exclusive events. Thus, equation (1) describes the probability of the event $E_k$ happening given that the event A has already happened. It does this by using the probabilities of A happening given each $E_j$ and the probabilities of each $E_j$ happening. (The denominator of the right hand side of (21) gives a standardizing

factor.) In other words, (21) gives the probability that the event A (which has occurred) was caused by $E_k$. That is to say, if A has happened and we know one of the $E_j$ also happened and we don't know which but we suppose it was $E_k$, then (21) gives the probability that our supposition is correct. We could do this for all $E_j$ if necessary and come up with a series of probabilities of each $E_j$ being the "cause" of A. We might consequently assume that A was caused by the $E_j$ with the highest (evaluated) probability.

How would we know $P[A|E_k]$? Simply from previous knowledge, and, as long as that knowledge is expressed in measurable frequencies, there is no controversy.

The Controversy

Suppose we have some data on failure rates of a device; call this data $\phi$. Suppose also we know that the data depends on some (unknown) parameter $\lambda$, the true failure rate. We can then use (21) to evaluate the probability of a certain value of $\lambda$ having "caused" the data. If we accept that $\lambda$ itself has a distribution then (21) can be rewritten.

$$f(\lambda|\phi) = \frac{f(\phi|\lambda)\ f(\lambda)}{\int f(\phi|\lambda)\ f(\lambda)\ d\lambda} \qquad (22)$$

$f(\phi|\lambda)$ is now the sampling distribution of the data given $\lambda$
$f(\lambda)$ is the prior distribution of $\lambda$, prior to data
$f(\lambda|\phi)$ is the posterior distribution of $\lambda$, posterior to (after) data

Now there is no problem defining $f(\phi|\lambda)$; we would assume or know the data to be exponential or whatever and use the corresponding formula in $\lambda$ to evaluate the probability of observing the data under a given $\lambda$.

e.g., for exponential data, if failure times $t_1 \ldots\ldots t_n$ were observed, then $\phi = (t_1, t_2 \ldots t_n)$ and

$$f(\phi|\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda t_i} \qquad (23)$$

$$= \lambda^n e^{-\lambda \sum_{i=1}^{n} t_i}$$

Where the controversy arises is in the concept of prior probability on a parameter (say $\lambda$). In classical statistics we use a confidence interval to define the uncertainty of our estimate of a parameter. The Bayesian approach is more direct: it assumes $\lambda$ is itself a random variable and then fits a distribution to that "variable." The classical statistician is only happy with the Bayesian approach if $\lambda$ is known to have a probability distribution measurable in the frequency sense. For example, if the failure data ($\phi$) were known to be either from Component A or Component B and we had to decide which, then using prior experience (measured by frequencies) of the failure rate exhibited by A and B, respectively, the classical statistician would be quite agreeable to use Bayes' theorem to decide which of the two components "caused" the data.

The Bayesian argues in the extreme that subjective probability is the only valid concept of probability and that data should only be used to enhance or refine.

Having accepted that a prior (based on subjective assessment) is reasonable, there is no further argument and the posterior drops out from (22). The posterior distribution is also a distribution of $\lambda$ and not a true sampling distribution. Bayesian distributions do, however, follow all the rules of distribution theory and integrate to unity.

To sum up, the posterior distribution combines prior knowledge with real data.

## The Likelihood Function

Likelihood is defined as the probability of observing a given set of data under some distribution that the data follows, hypothesised or known. Likelihood is written $L(\lambda|\phi)$ and defines the likelihood of the data coming from a distribution with parameter $\lambda$. Thus, a maximum likelihood estimator (MLE) is the one which maximizes the likelihood function and may be derived using differential calculus; quite simply, this is not done here.

For a set of data $\phi$ as before, $(\phi = t_1,....t_n)$

$$L(\lambda|\phi) = \prod_{i=1}^{n} f(t_i)$$

so if f is exponential

$$L(\lambda|\phi) = \prod_{i=1}^{n} \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} t_i}$$

so the MLE is found by setting $\frac{dL(\lambda|\phi)}{d\lambda} = 0$ and solving for $\lambda$.

It can now easily be seen that the likelihood function is the same as $f(\lambda|\phi)$ in (22). And since the denominator of (22) is a standardizing term, we can immediately restate Bayes' theorem (for the third time) as

$$\text{Posterior} \propto \text{likelihood} \times \text{prior} \qquad (24)$$

Note that the difference between the likelihood function $L(\lambda|\phi)$ and the sampling distribution $f(\phi|\lambda)$ is that L is a function of $\lambda$ given $\phi$; and f is a function of $\phi$ given $\lambda$.

If necessary, (24) could be expressed mathematically and the proportionality replaced by an equality by introducing a standardized likelihood function (by dividing by an integration of L over $\lambda$).

Numerical Example

The following example falls into the controversial category of Bayesian application.

One day in 1981, Joe Steinkirchner asked thirty people to estimate the number of hours of use of a certain piece of equipment (which they all used) per year. They said "160, 186, 390, 466, 560, 762, 1116, 1176, 1416, 1480, 1788, 1800, 2040, 2140, 2416, 2496, 2520, 2608, 2700, 2788, 2808, 2904, 2960, 3708, 3744, 3960, 5040, 5120, 5688, 7520." Joe didn't know who to believe, so he went and collected the following data:

860.9 1069.6

His problem was to combine all the expert opinion data (clearly of value, he argued) with the data (far too sparse to be used alone but also of value).

The solution was defined as follows:

The opinion data is found to be reasonably normal (by Kolmogorov-Smirnov test) with mean 2482 and standard deviation 1740.5. If we accept that the opinion data is a valid prior, then the prior is given by

$$f(\mu) = \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp - \frac{1}{2}\left[\left(\frac{\mu - \mu_0}{\sigma_0}\right)^2\right] \qquad (25)$$

where

$\mu$ is the mean number of hours of use (remember that $\mu$ is here considered a random variable)

$\mu_0$ is the mean of the opinion data

$\sigma_0$ is the standard deviation of the opinion data

Hence

$$\mu_0 = 2482$$
$$\sigma_0 = 1740.5$$

Next we have to establish the sampling distribution of the data: $f(\phi|\mu)$, where $\phi$ is the data. Clearly there is not enough data to derive a distribution but we glibly assume it to be normal (which may be justifiable) and even more glibly estimate the mean and standard deviation from the data to illustrate the technique. Ideally, though, we'd collect more data for better estimates or at least know the sampling distribution.

Hence

$$\bar{x} = 965.3$$
$$\sigma_1 = 147.6$$

So

$$L(\mu|\phi) = \left(\frac{1}{\sqrt{2\pi}\,\sigma_1}\right)^n \exp\left[-\frac{1}{2\sigma_1^2}\sum (x_i - \mu)^2\right] \tag{26}$$

where

$x_i$ are the individual data points
n = number of points = 2

It may be shown (Ref. 9) that this is the same as

$$L(\mu|\phi) \propto \exp\left[-\frac{1}{2}\left(\frac{\mu - \bar{x}}{\sigma_1/\sqrt{n}}\right)^2\right] \tag{27}$$

which is intuitively reasonable (where $\bar{x}$ is the mean of the data and $\propto$ indicates proportionality).

So (22) and (23), give

$$f(\mu|\phi) = \frac{\frac{1}{\sqrt{2\pi}\,\sigma_o}\exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_o}{\sigma_o}\right)^2\right]\exp\left[-\frac{1}{2}\left(\frac{\mu-\bar{x}}{\sigma_1}\right)^2\right]}{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\,\sigma_o}\exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_o}{\sigma_o}\right)^2\right]\exp\left[-\frac{1}{2}\left(\frac{\mu-\bar{x}}{\sigma_1}\right)^2\right]d\mu} \quad (28)$$

where $f(\mu|\phi)$ is the posterior distribution of $\mu$.

The math is clearly going to be tedious; fortunately, Box has already done it (Ref. 9) and found that

$f(\mu|\phi)$ is normal with mean $\frac{1}{1/\sigma_o^2 + n/\sigma_1^2}\left[\frac{\mu_o}{\sigma_o^2} + \frac{n\bar{x}}{\sigma_1^2}\right]$

and variance

$1/(1/\sigma_o^2 + n/\sigma_1^2)$

And, since before we had

$\mu_o = 2462$
$\sigma_o = 1740.5$
$\bar{x} = 965.3$
$\sigma_1 = 147.6$
$n = 2$

Substitution in the above formulae gives the posterior mean as 970.7 and the posterior standard deviation as 104.2.

Hence our best estimate of the number of days' usage is 970.7 by combining prior information with data. Pictorially, the example is depicted in Figure 20.

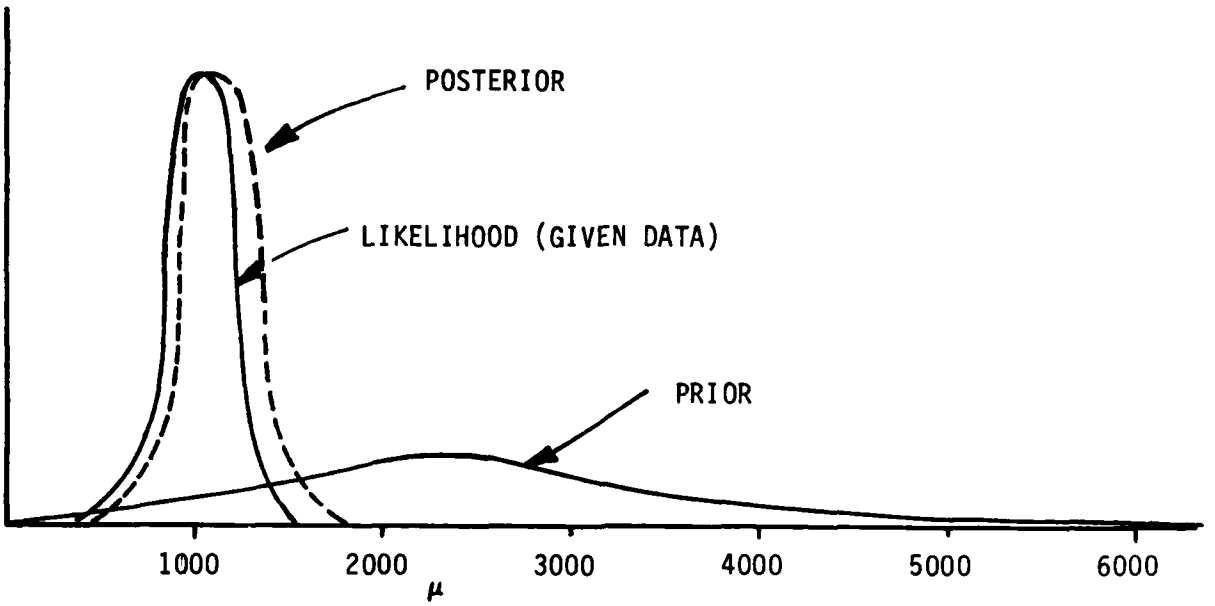



FIGURE 20: PRIOR AND POSTERIOR DENSITIES

It is interesting to note that the posterior is closer to the data (likelihood) than it is to the prior because of the relatively large prior variance. Of course, the opposite effect can occur, depending on the data and the prior in each case.

One would be unlikely to use this technique with only two data points, but the example chosen keeps the numbers simple to illustrate.

Sequential applications of Bayesian methods can be used as more data becomes available, using the posterior at one iteration as the prior for the next.

## Conclusions

If one accepts the use of a prior, without being able to measure it except subjectively, there are vast amounts of work which can be carried out using Bayesian methods (Ref. 9). So far, the main applications have been in decision theory, using Bayes' theorem to decide which of a

number of options is the "best" in the face of multiple risks. In reliability this has been used in choosing which of two MTBFs an equipment exhibits (Refs. 8 and 32).

An example has been given to suggest how data and opinion might be combined mathematically. It is an alternative to using the ubiquitous "engineering judgement" phrase and must surely lend more credence to technical reports.

The mathematics are quite involved but solutions exist in text books for various combinations of prior and sampling distributions.

Even if we reject the use of a subjective prior, we ought to make use of measureable priors whenever they are available.

# CHAPTER 8

## Graphical Regression Analysis

CHAPTER 8

# Graphical Regression Analysis

## Introduction

This chapter gives a means of fitting linear regression line without a formal analysis. This is useful as a quick means of trying out regressions. The accuracy of the method is shown by comparing to Figure 21, which gives formal analysis of the same data.

## Procedure

(i) Plot the data shown in Figure 21 on the usual X, Y plane.

(ii) Discard any outliers.

(iii) Compute the average X and $\bar{x}$ and the average Y and $\bar{y}$ and plot the point $(\bar{x}, \bar{y})$. This is shown as $\theta$ in Figure 21. (If there are a lot of points, the medians may be used to save time.)

(iv) Construct a perpendicular onto the X axis, through $\theta$.

(v) Draw the line to the left of the perpendicular and though $\theta$, which divides the points to the left of the perpendicular in half. This line is $A\theta$ in the figure.

(vi) Do the same on the right hand side of the perpendicular ($\theta B$). Extend this line to the left of the perpendicular to end up with the line BC.

(vii) Bisect the angle $A\hat{\theta}C$. This is an approximation to the least squares line of best fit.

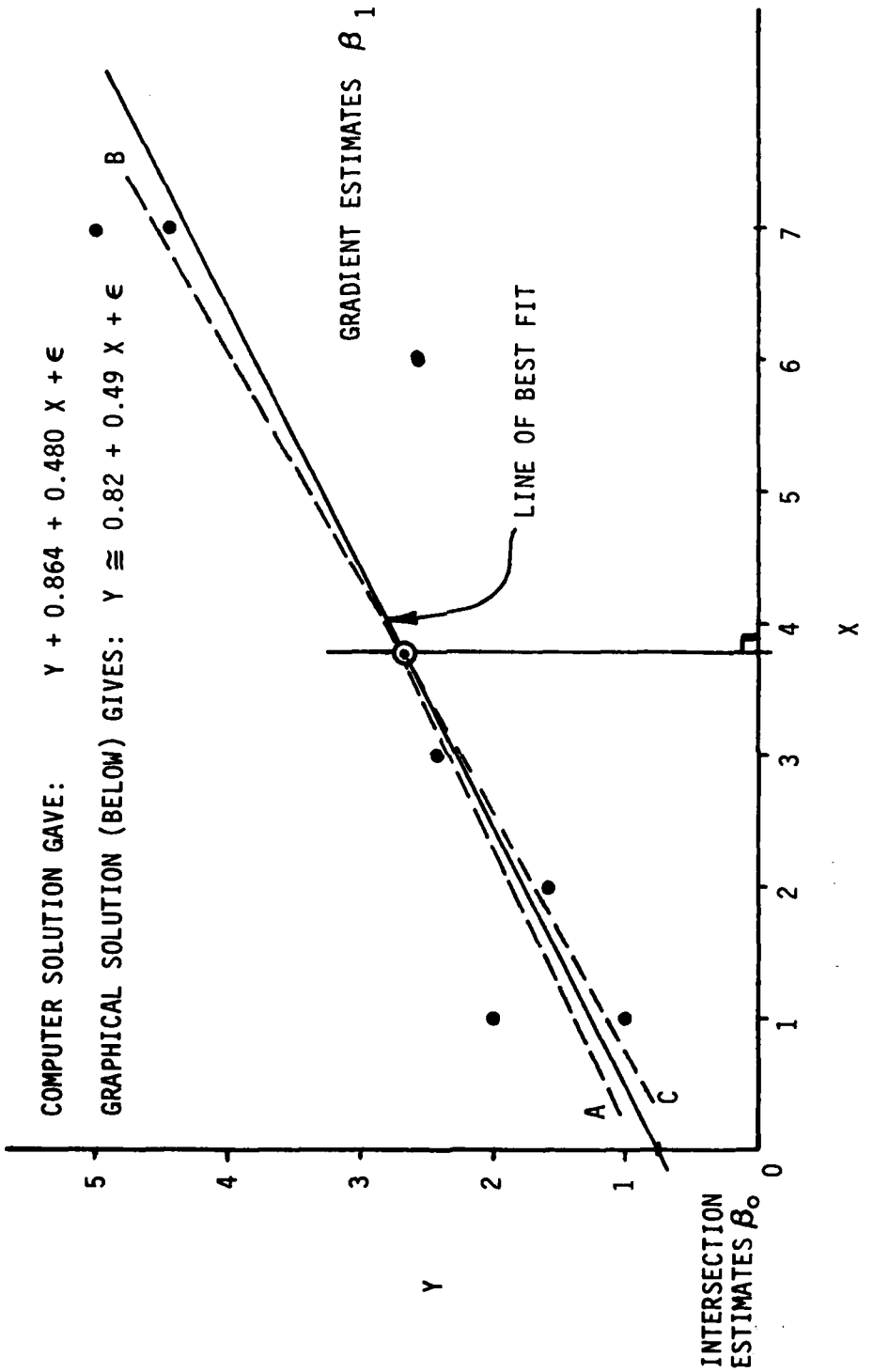



FIGURE 21: GRAPHICAL APPROXIMATION TO A LEAST SQUARES SOLUTION

(viii) If the equation of the line is required in the form $Y = \beta_0 + \beta_1 X$, then the values of $\beta_0 \beta_1$ are easily found by intercept and slope as shown. In the example, this equation is given by $Y = 0.82 + 0.49X + \epsilon$. This compares well to the formal least squares solution (see Figure 21).

Restrictions

(i) Though the method is quite accurate, no confidence intervals on the coefficients are given. The original paper (Ref. 22) included a method for graphical interval estimation.

(ii) The method will only fit a linear relation, though others are possible by transforming variables.

(iii) The method should eventually be backed up by formal analysis, and goodness-of-fit tests are explained in Chapter 9.

Uses

The method may be used wherever a relation between two variables is suspected. It could be used to remove subjectivity from Weibull plots but a minimax fit would seem truer to the sense of a Weibull (or other probability) plot. A quick minimax fit is given in Ref. 20, which also gives a useful review of graphical line fitting.

Transformations

Examples of useful transformations are shown in Figure 22. These figures give the original (non-linear) data together with the linearizing transformation and the resultant linear relation.

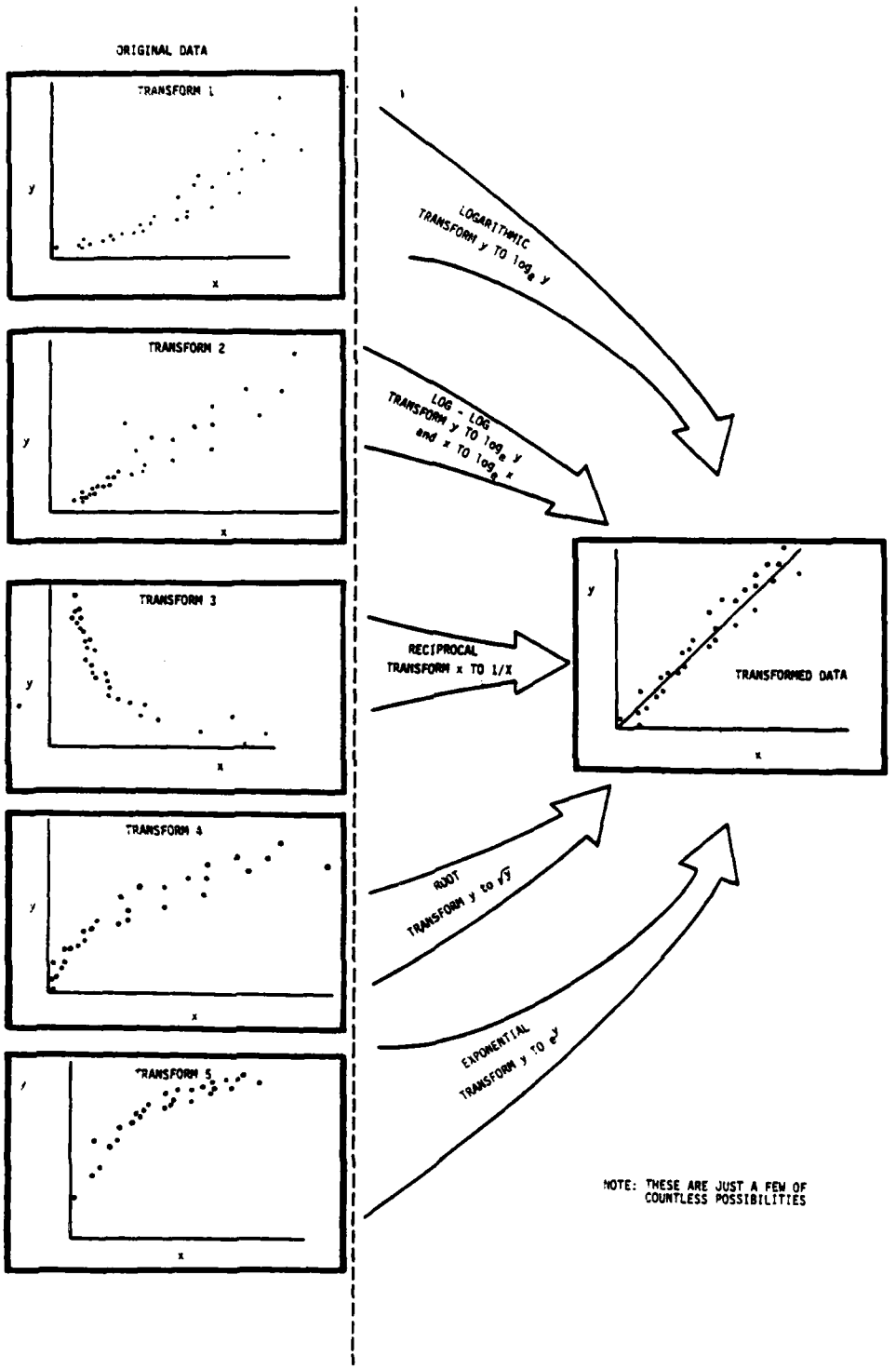



FIGURE 22: EXAMPLES OF DATA TRANSFORMATIONS

# CHAPTER 9

## Regression Analysis and Least Squares Theory

CHAPTER 9

# Regression Analysis and Least Squares Theory


## Summary

This note attempts to combine some useful ideas into a practical "user guide" to regression analysis for the non-specialist. The least squares approach is used. The purpose of regression and the differences between predictive and scientific models are highlighted. Some common errors of technique and interpretation are underlined. Some recent numerical examples of regression analysis in IITRI/RAC projects are presented to illustrate the power of the method over and above its basic "number crunching" uses.


## Introduction

Regression analysis using least squares theory is just one way of establishing a relation between a number of variables, one of which is dependent on the rest. For example, such an analysis might be used to establish a relation between electrostatic sensitivity of a component and the resistance and capacitance of the test circuit employed. Note the parallel with Chapter 2, except that here the variables are continuous (rather than discrete).

When there are only two variables involved, (e.g., sensitivity and capacitance), it is reasonably easy to see a line of best fit and we only need regression to optimize it. When a third variable is introduced (e.g., resistance), it can be difficult to see a fit or guess its form. When ten further variables are introduced, the brain becomes hopelessly confused and regression analysis is essential. Of course there may be some law of physics available for all the variables (or we could try to establish one from first principles) but the law would

still need to be tested against the data and, again, regression analysis is far quicker, more accurate and less biased than trying to draw ten graphs and look at them all together. In addition, the regression tells us how "good" the fit is and answers the question "How good is good?"

Regression-type methods can be applied to many other problems, too, including prediction models, optimization problems, and control. This note only looks at least squares solutions though other methods are available (e.g., minimax, non-parametric and Bayesian).

## The Linear Model

A model of the form

$$Y = \beta_0 + \beta_1 X + \epsilon \tag{29}$$

is referred to as the linear model, where Y is a dependent variable, X is the independent variable, $\epsilon$ is experimental error (or noise) and the $\beta$s are constants. Thus Y depends on X. Strictly, X should be measured without error so that $\epsilon$ describes the error in Y alone. The best way to do this is to have X as a known variable which can be controlled (e.g., X is a particular number of pulses; integer numbers are without error). In practice, error is often present in X; it is difficult to meet the full theoretical requirements.

The term "linear" refers to the coefficients. The highest power of X is termed the order. Hence (29) is a first-order linear model. An example of a second-order linear model would be

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon \tag{30}$$

An example of a non-linear model would be

$$Y = \beta_0 + X^{\beta_1} + \epsilon \tag{31}$$

Non-linear models are difficult to solve and so we usually seek a linear model or one which is intrinsically linear (i.e., transformed to linear by a transformation such as taking logarithms).

An intrinsically linear example would be

$$Y = e^{\beta_0} X^{\beta_1} \epsilon \tag{32}$$

since then

$$\log Y = \beta_0 + \beta_1 \log X + \log \epsilon \tag{33}$$

which is linear. (33) can then be solved as if it were a linear model and transformed back to the form of (32) after solution (though some bias may be introduced unless the log $\epsilon$ are normally distributed).

Why the Linear Model?

Many problems in nature follow an additive law which culminates in the observed normal distribution. Examples are plant and animal growth, where factors such as food, vitamins and heredity have an additive effect on height of the final plant or animal (i.e., each factor influences height by a certain positive or negative amount).

Many problems in physics and engineering, however, follow (at least to a first-order approximation) a multiplicative law which culminates in the observed lognormal distribution. Examples include reliability where factors such as stress, temperature, vibration, etc., have a multiplicative effect on failure rate (i.e., each factor increases the failure rate by a certain proportion).

Mathematically, the additive model is given by (29) above, the multiplicative by (32) above. Since we have already seen that (32) is

transformable to (29), solutions of the additive and multiplicative laws of nature and physics/engineering are possible by regression analysis.

Solution of the Linear Model - Least Squares Theory

Consider (29) above as applied to the $i^{th}$ observation of a set of n data points.

$$\text{Then } Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{34}$$

$$\text{Rewriting } \epsilon_i = (Y_i - \beta_0 - \beta_1 X_i) \tag{35}$$

Then to take care of any negative results, consider

$$\epsilon_i^2 = (Y_i - \beta_0 - \beta_1 X_i)^2 \tag{36}$$

Then a measure of total experimental error is given by

$$\sum_{i=1}^{n} \epsilon_i^2$$

Clearly, the best model will explain most of the variation in the data and leave little to experimental errors. Hence, we need to minimize

$$\sum_{i=1}^{n} \epsilon_i^2 \quad \text{such that} \quad \sum_{i=1}^{n} \epsilon_i = 0$$

and hence seek solutions to

$$\frac{d}{d\beta_0} \left[ \sum_{i=1}^{n} \epsilon_i^2 \right] = 0 \tag{37}$$

$$\frac{d}{d\beta_1} \left[ \sum_{i=1}^{n} \epsilon_i^2 \right] = 0 \tag{38}$$

(37) and (38) are generally referred to as the normal equations. A sketch of the situation is given below to clarify, and it is seen that we seek to minimize the squared deviations of the data from our fitted line (model); hence the term least squares. This concept is illustrated in Figure 23.

The sum of squared deviations $\Sigma\epsilon_i^2$ is called the residual sum of squares and is analogous to noise on an oscilloscope. If $b_0$ and $b_1$ are the solutions for $\beta_0$ and $\beta_1$ in (37) and (38) then

$$b_0 = \bar{Y} - b_1 \bar{X} \tag{39}$$

$$b_1 = \Sigma(X_i - \bar{X})\ (Y_i - \bar{Y}) / \Sigma(X_i - \bar{X})^2 \tag{40}$$

where $\bar{X}$ and $\bar{Y}$ are the means.

Similar arguments can be applied to higher-order models or models with more than one independent (X) variable. Thus, (39) and (40) give a means of putting the (linear) line of best fit through a set of data.

How Good is the Fit?

It is obvious that individual data points will deviate from the overall mean. Most of that deviation should, ideally, be explained by the regression equation. So the deviation $\epsilon_i$ of the $i^{th}$ data point from the regression lines should be small compared to the deviation of the regression line $(\hat{Y}_i - \bar{Y})$ from the overall mean (where $\hat{Y}_i$ is the estimated value of Y from the regression line).

The multiple coefficient of determination, usually denoted by $R^2$, evaluates this for all the data together, as:

$$R^2 = \frac{\Sigma(\hat{Y}_i - \bar{Y})^2}{\Sigma(Y_i - \bar{Y})^2}$$

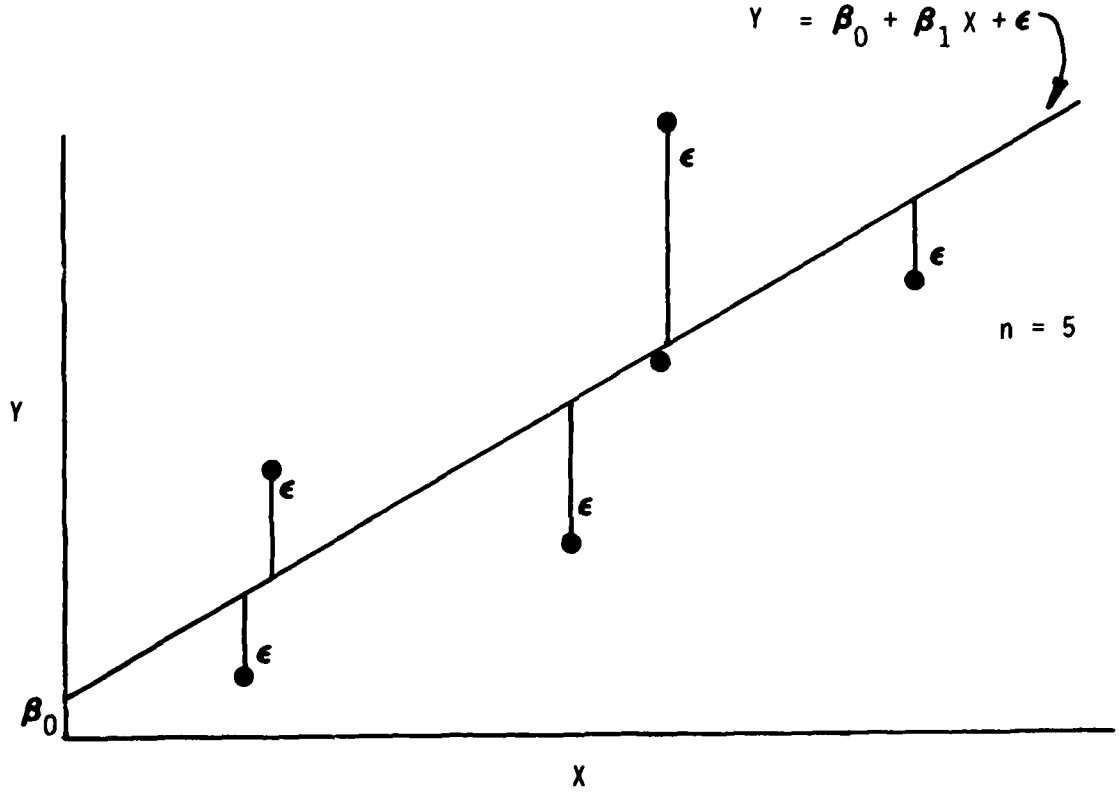



FIGURE 23: MINIMIZING THE SQUARED DEVIATIONS

Ideally, $R^2 = 1$ and all the points lie on the fitted line. In practice, $R^2$ is between 0 and 1. $R^2 \times 100$ is also used to express $R^2$ as % and is then called "percent of total variation explained." Figure 24(a) and (b) will clarify.

There is no objective means of assessing what value of $R^2$ suggests a good fit other than "close to 1." $R^2$ is, however, a useful clue in the search for the line of best fit. Note that so far there have been no assumptions as to the form of distribution of the data.

For two variables, R is simply the correlation r described in Chapter 1.

How Good is Good?

If it is possible to assume that the $\epsilon_i$ are normally distributed with mean zero and that the autocorrelation of $\epsilon_i$ is 0, then further evaluation of the line is possible.

By comparing the predicted values of $\hat{Y}$ (i.e., $Y_i$) to the experimental errors ($\epsilon_i$) and allowing for the number of data points fitted (n), we derive an F statistic with 1 and n-2 degrees of freedom. F measures objectively how well the line fits the data. It is the same statistic as was used in analysis of variance (Chapter 2) and for the same reasons. Crudely, the larger the F the better the fit. The F value may be evaluated properly by tables using 1 and (n-2) degrees of freedom.

This chapter assumes that the variance is equal (homoscedastic) for all values of the independent variable(s). If this does not hold, a weighted regression analysis should be used (Ref. 23).

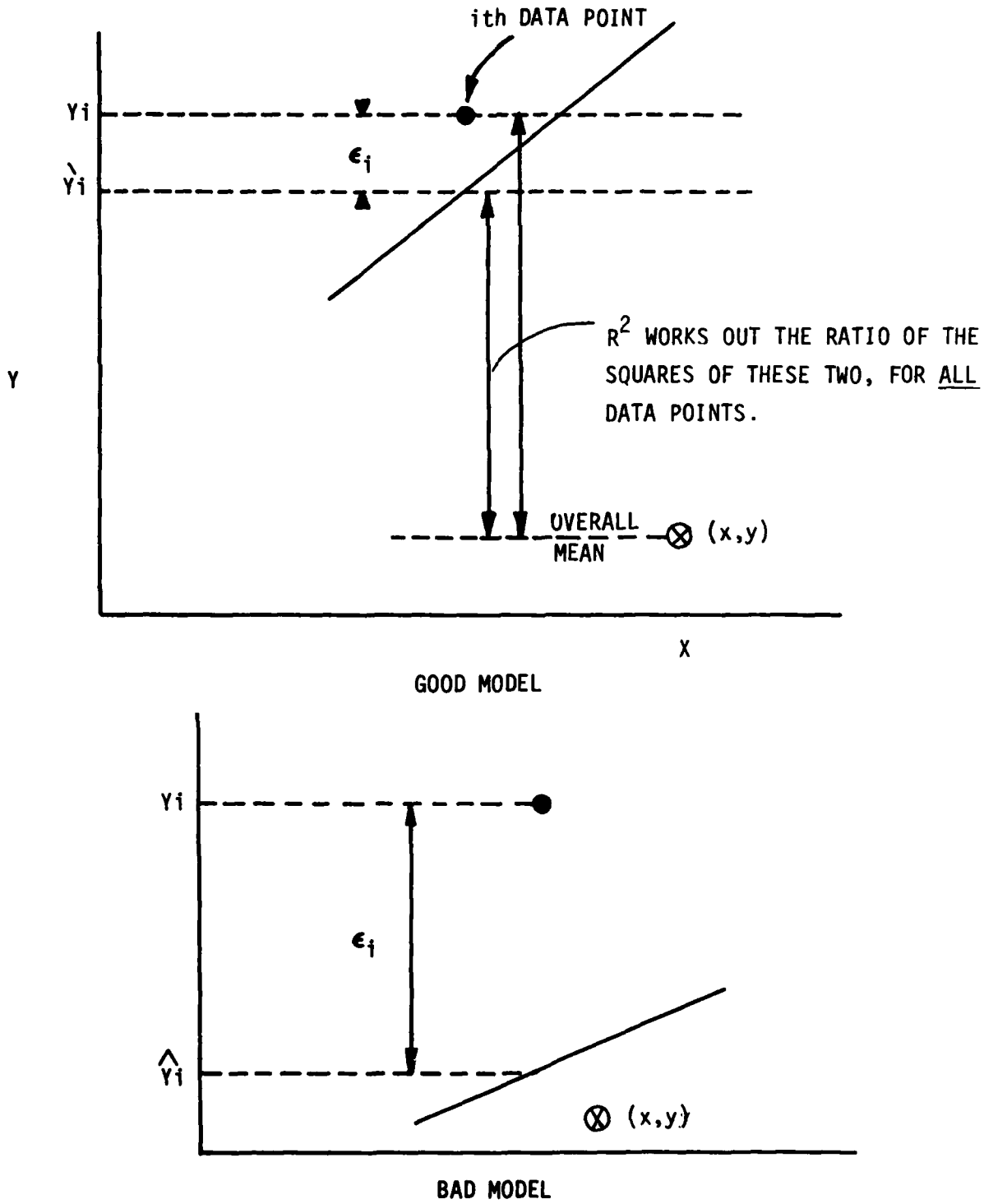



FIGURE 24: EFFECT OF GOOD AND BAD MODELS ON RESIDUAL VARIABILITY AND $R^2$

Is There a Better Fit?

It may be that a model of the form of (30) will give a better fit to the data by including a term in $X^2$. If this were suspected then the normal equations would be solved again, this time for $\beta_0$, $\beta_1$ and $\beta_2$. To compare this fit to the first fit (that of equation (29)), compare the respective $R^2$ and F values.

It is also possible to evaluate F at each stage of fitting the model, i.e., for just

$Y = \beta_0 + \epsilon$ first
then $Y = \beta_0 + \beta_1 X + \epsilon$
and then $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$.

Other statistics are available to evaluate what you gained at each stage. Computer programs do all this automatically and are called stepwise regression solutions. Such programs only add the next stage if it produces a significant improvement in the model as measured by F.

Clues as to whether an $X^2$ term or an $X^3$ or an $e^X$ or whatever should be added may be seen in a plot of the residuals against X. Knowledge of the scientific relation between the variables is invaluable and should be sought also.

It is not wise to try to get too good a fit. By fitting more and more variables it is easy to get an $R^2$ of 0.99. However, this may not fit a second set of data and we would merely be modelling the noise! In practice some noise is always present. The following example shows how the correct amount of noise is sometimes able to be calculated. In other words, we should not try too hard and should know (statistically) when to stop. This explains why $R^2$ alone is not a good assessment of a model. It is very rare that more than 4 or 5 variables are appropriate and 8 is probably a maximum.

Numerical Example

Using the following data, the principle of regression analysis is illustrated. Here, n = 21.

TABLE 21: REGRESSION DATA

| X | Y | X | Y | X | Y | X | Y |
|---|---|---|---|---|---|---|---|
| 1 | 0.4 | 2 | 1.2 | 4 | 2.0 | 5 | 2.8 |
| 1 | 0.8 | 2 | 1.2 | 4 | 2.4 | 5 | 3.0 |
| 1 | 1.0 | 2 | 1.4 | 4 | 2.6 | 5 | 3.2 |
| 1 | 1.2 | 2 | 1.6 | 4 | 2.8 | 5 | 3.4 |
| 1 | 1.6 | 2 | 1.8 | 4 | 3.0 | 5 | 3.6 |
| | | 2 | 2.4 | | | | |

using the model (29), i.e.,

$$Y = \beta_0 + \beta_1 + \epsilon$$

and solving for $\beta_0$, $\beta_1$ from (39) and (40) gives

$$b_0 = 0.487$$
$$b_1 = 0.535$$

So the model is given by

$$Y = 0.487 + 0.535\ X + \epsilon$$

The model and data is sketched in Figure 25.

Since the total sum of squared deviations about the grand mean is evaluated at 17.466 and the residual sum of squares (otherwise called $\Sigma\epsilon_i^2$ or simply noise) is found to be 2.875, then

$$R^2 = \frac{17.466 - 2.875}{17.466} = 0.856$$

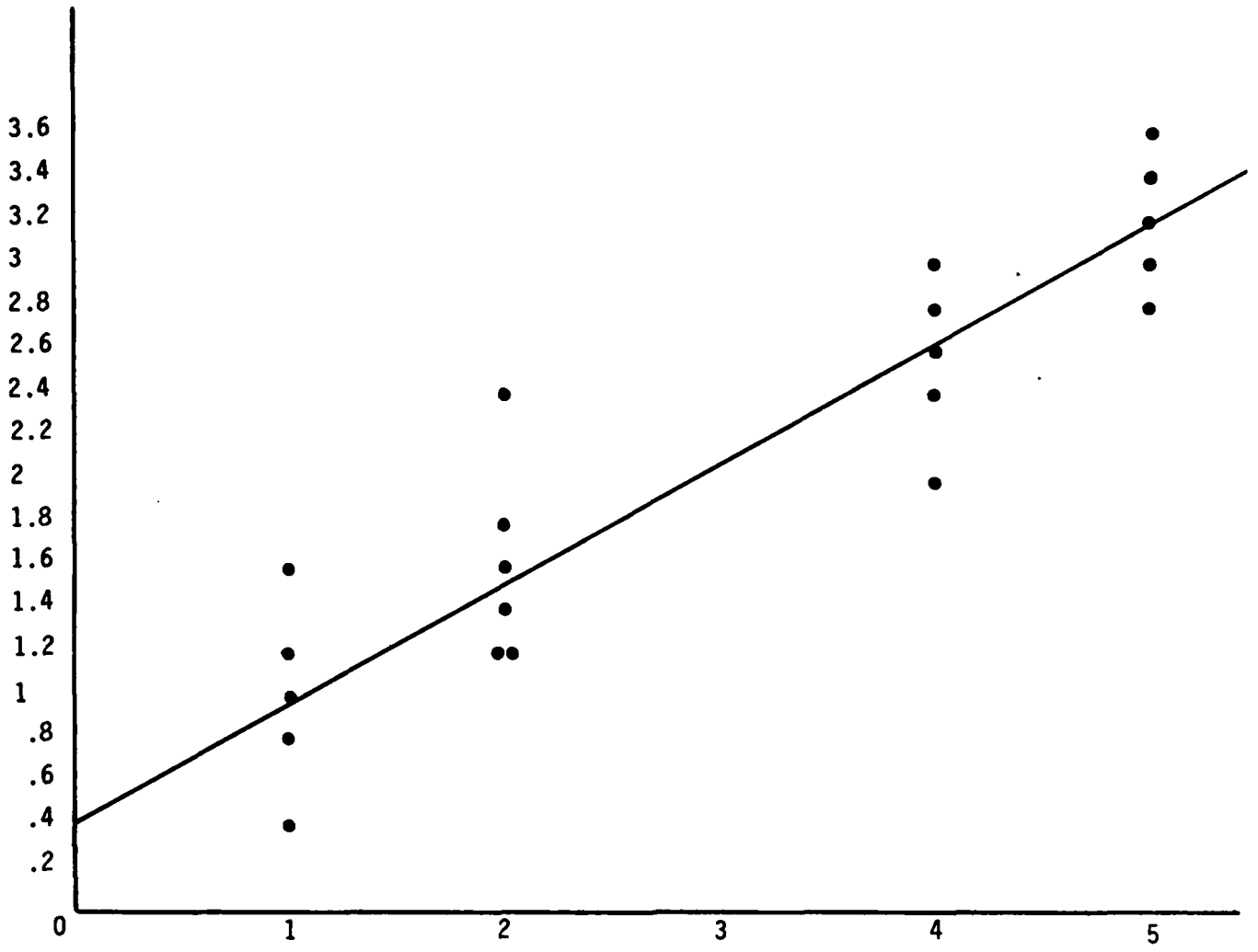



FIGURE 25: REGRESSION OF Y ON X

i.e., 85.6% of the variation in the data is accounted for by the model. This indicates a satisfactory model so far.

A residual plot for $\epsilon_i$ (i.e., $Y_i - \hat{Y}_i$) is sketched in Figure 26.

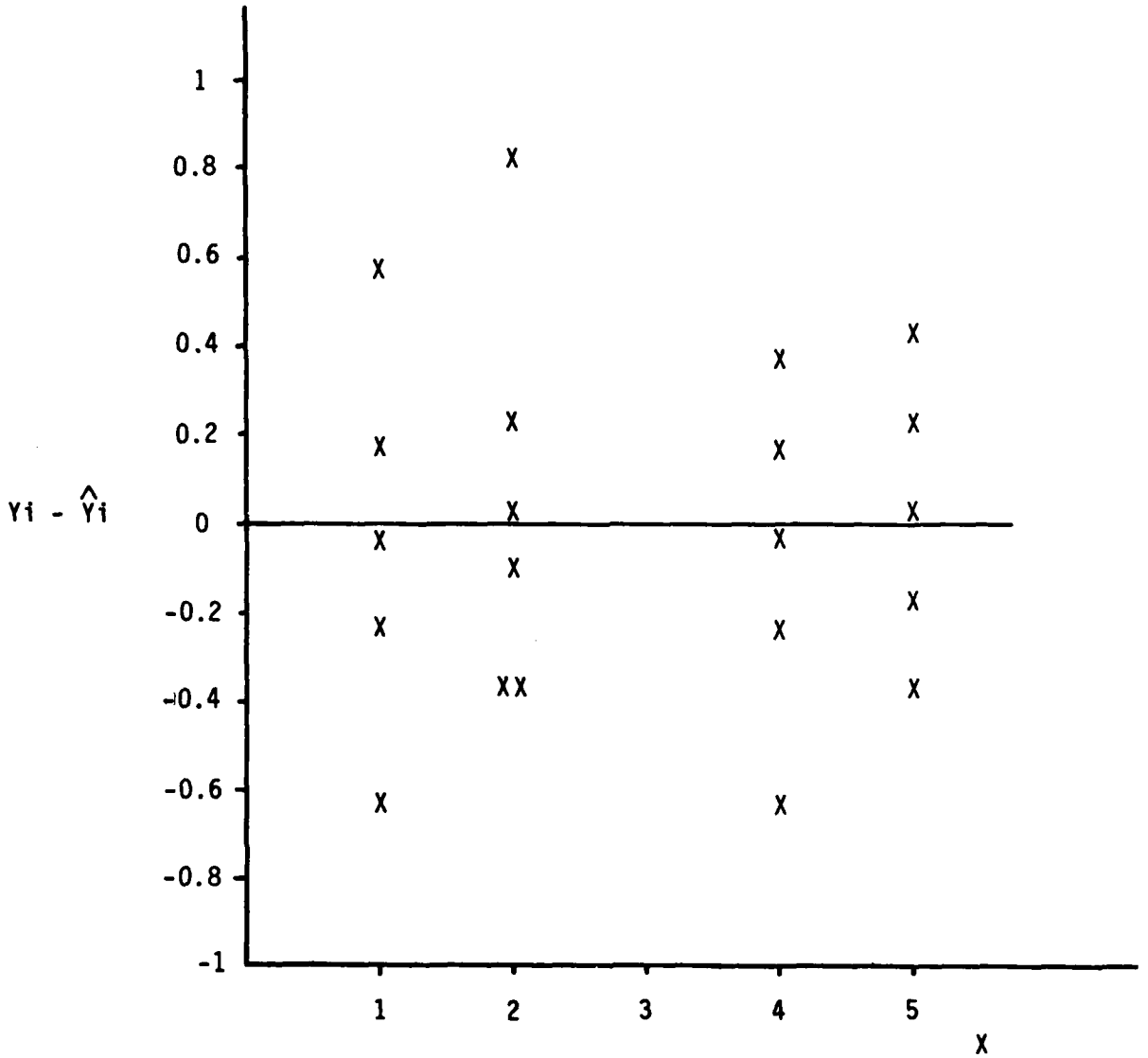



FIGURE 26: RESIDUAL PLOT

The plot does not show any obvious trends, with the data randomly scattered about the zero line as required. (In fact, statistical tests of randomness are available for a proper analysis, Ref. 23.)

The data seem to be reasonably symmetric about the regression line, so we assume normality and can now incorporate the F statistic.

This is done in a standard table (explained fully in Chapter 2).

TABLE 22: REGRESSION STATISTICS

| Source of Variation | Sum of Squared Deviations | Degrees of Freedom | MS | F |
|---|---|---|---|---|
| Due to Regression | 14.591 | 1 | 14.591 | 96.54 |
| Residuals (Noise) | 2.875 | 19 | 0.151 | |
| Total, about mean | 17.466 | 20 | | |

The degrees of freedom due to the regression is equal to the number of independent variables, the total degrees of freedom is (n-1) and the residual degrees of freedom follow by subtraction. (F is found by dividing the regression mean square by the residual.)

We find that the F value is quite high and consulting statistical tables (Ref. 3) shows the critical F with 1, 19, d.f. at the 5% level is 4.38. At the 0.1% level, the critical value is 15.1. Hence, the regression is very highly significant. We can therefore be confident that the fitted line follows some scientific law rather than just noise.

It is also possible to go even further if more than one observation is available at each value of X. Then we can actually evaluate how close our model is to the perfect model by extracting the experimental error, which may be estimated from a series of observations of Y at the same X value.

We do this simply by adding the squared deviations of each Y from the mean Y at <u>each</u> value of X. Then by subtraction from the (previously evaluated) residual sum of squares we find the lack of fit. The following more complex table clarifies.

TABLE 23: REFINED REGRESSION STATISTICS

| Source | SSD | df | MS | F |
|---|---|---|---|---|
| Due to Regression | 14.591 | 1 | 14.951 | |
| Due to lack of fit | 0.043 | 2 | 0.0215 | 0.129 |
| Pure error | 2.832 | 17 | 0.1666 | |
| Total (about mean) | 17.466 | 20 | | |

Comparing the lack of fit to the pure error gives an F of 0.129, which is not significant. Thus there is very little lack of fit and no need to try to improve the model. (Without repeat observations at given X, it is not possible to do this and assess the lack of fit.)

Computer Solution

The whole problem is much easier with the aid of a machine though blind analysis can be dangerous. Additional useful features are given usually, including a printout of residuals and % errors for each point, as well as including standard errors on each estimate. This is most important and should be included in any report on (or use of) regression analysis.

Scientific and Predictive Models

These are often confused and interchanged on the assumption that they are the same; they are not.

- o Scientific models - establish a law between a number of variables from an understanding of their interaction. Typically

they are derived from first principles. An example would be Boyle's law, which says that the pressure in a vessel times its volume is constant:

$$PV = K \tag{41}$$

A good scientific model holds for the full range of P and V, and its use is therefore unrestricted by range. Their use is universal, subject to the constraints of underlying assumptions, and they can be used for prediction except that they do not describe the experimental error.

o Predictive models - established from empirically observed relations between variables by regression analysis (or similar method). A major difference between these and scientific models is that they include a stochastic variable, i.e., one which models the experimental error. Thus an ideal predictive model for pressure might be

$$P = \frac{K}{V} + \epsilon \tag{42}$$

(Though a predictive model will not generally follow the form of a scientific one.)

For a very complex situation involving hundreds of variables we might find that just a few of them could be used for predictive purposes. For example, reliability is influenced by many parameters but a reasonable fit for failure rate under given conditions might be given by a simple model like

$$\log \lambda = \beta_0 + \log (T_1) + \log (T_2) + \epsilon \tag{43}$$

where

$\lambda$ is failure rate

$T_1$ and $T_2$ are junction and ambient temperatures

The derived predictive model may or may not be a scientific model; usually it won't be. A predictive model only holds over a certain range of the independent variables which made up the (analyzed) data. (Extrapolation requires caution.) However, it is also required that the form of the predictive model make physical sense.

A predictive model of Y on X may not be used to predict X from Y! Nor is it valid to fiddle around with the model to make claims as to the relation between two of the independent variables nor to change the subject of the formula to find one independent variable in terms of the dependent and the other independent variables. Predictive models should never be quoted without confidence intervals on the coefficients ($\beta$s) since to do so would ignore the stochastic element, which is crucial. Construction of intervals is straightforward, particularly when using a computer solution (Ref. 3).

Regression analysis may be used to help derive a scientific model by careful construction and testing, as indicated in the example given below.

The Multiplicative Model

Recall (32): $Y = e^{\beta_0} X_1^{\beta_1} \epsilon$

which may be solved by taking logs and using a least squares approach. If the F-statistic is to be used then it is necessary that log $\epsilon$ is normally distributed. This is often found to be the case in reliability studies to a reasonable approximation. Having solved the transformed model (33) and checked $R^2$ and F values and found a good fit, the final step is to re-transform back again by taking exponents.

Not all transformations preserve the statistical properties; specifically, if the expected value of the transformed data is not equal

to the transform of the expected value then things like confidence intervals become approximate.

Example

The multiplicative model can be very useful in investigating engineering problems. For example, a recent ESD analysis had 26 data points on voltage of failure (V) together with resistance (R) and capacitance (C) of the test circuit. Physical/electronic knowledge suggested that V was proportional to R but inversely proportional to $\sqrt{C}$. So we fitted a multiplicative model given by:

$$V = \alpha R^{\beta_1} C^{\beta_2} \epsilon$$

where $\alpha$, $\beta_1$ and $\beta_2$ are constants. The solution was (with standard errors included)

$$\alpha = 4.692$$
$$\beta_1 = 0.647 \pm 0.0584$$
$$\beta_2 = -0.408 \pm 0.0519$$

suggesting that the inverse $\sqrt{C}$ law was feasible. This solution when plotted out gave us valuable clues and led to a very good fit with a modified model:

$$V = -158.1 + 0.658R + \frac{6966.5}{\sqrt{C}} + \epsilon$$

The fit of such a model may be appreciated by comparing graphically and statistically to the multiplicative model and an additive one. If we define these three models as Types A, B, and C, their form and fit may be summarized as follows:

TABLE 24: COMPARISON OF THREE REGRESSION MODELS

| Model | Description | Form | $R^2$ | $F_R$ | $F_C$ |
|---|---|---|---|---|---|
| Type A | Multiplicative | $V = \alpha R^{\beta_1} C^{\beta_2} \epsilon$ | 87.5% | 122.7 | 61.8 |
| Type B | Additive, with C in logarithmic transform | $V = \alpha' + \beta_1' R + \beta_2' \log C + \epsilon$ | 90.6% | 194.9 | 70.2 |
| Type C | Final model | $V = \alpha'' + \beta_1'' R + \frac{\beta_2''}{\sqrt{C}} + \epsilon$ | 90.8% | 198.6 | 72.3 |

where $R^2$ is the usual coefficient of determination. $F_R$ and $F_C$ are the F-statistic for each variable. Clearly all F values are very highly significant (critical $F \simeq 14.2$ at 0.1% significance) but the Type C model gives the highest. Since its $R^2$ value is also superior, the Type C model is superior on all three measures. The three models are sketched in Figures 27 and 28 for graphical comparison and it may be seen that the Type B and C models give the best fit, though it is hard to chose between them without statistical measures.

In reliability problems, some multiplicative model is highly recommended in looking for a valid prediction model, though the penalty of stopping with the simple multiplicative model may be high, as was the case with this example.

## Data Problems

The validity of a regression analysis is clearly dependent on the quality of the initial data. Every attempt should be made to collect equal amounts of data over the combined range of independent variables. Experimental designs should ideally be used but these are often impractical in RAC studies. Statisticians always try to achieve balanced data sets, and mathematicians talk of orthogonality. It is advisable to see how this talk affects your study before collecting data or doing a series of tests.

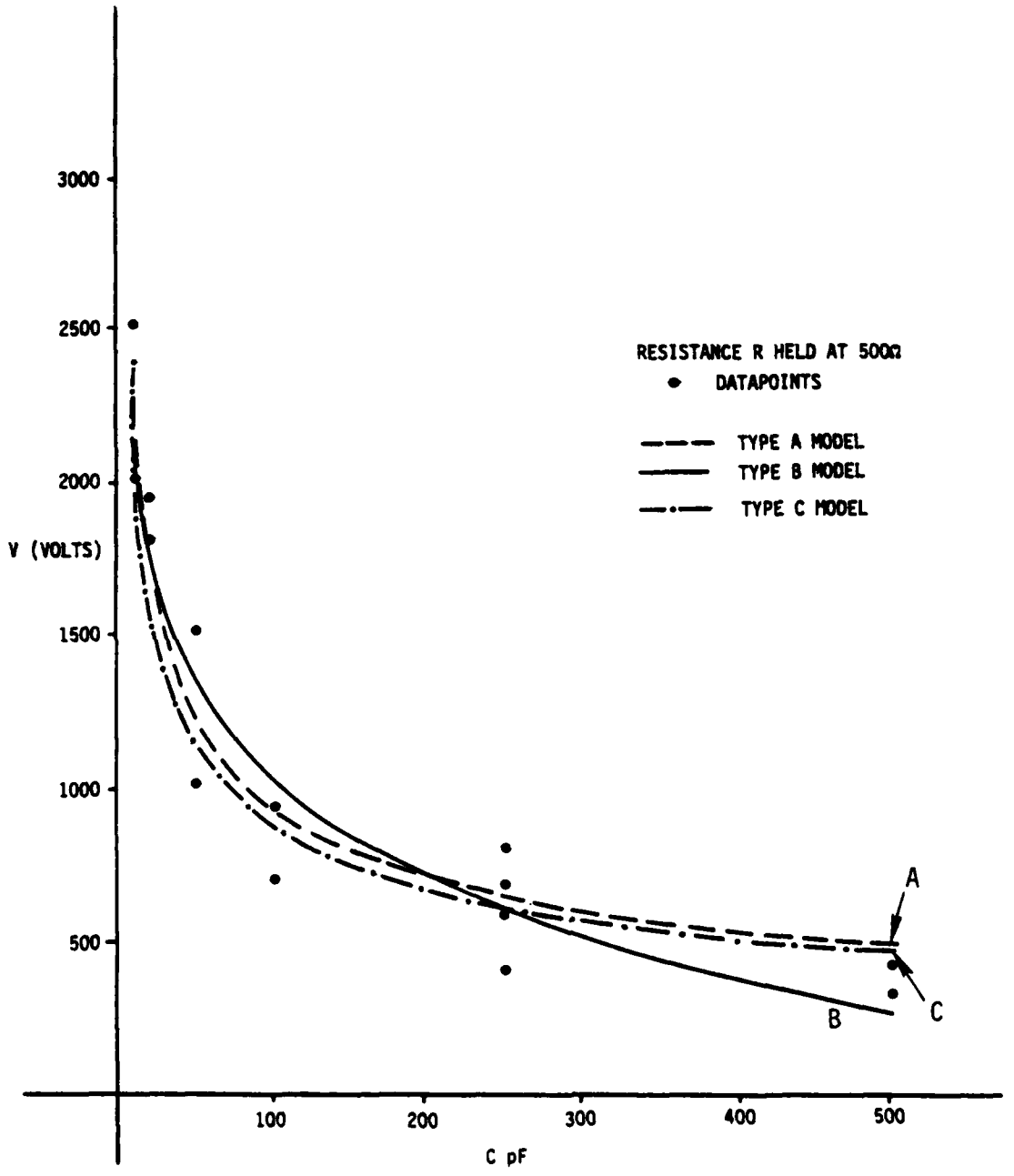



FIGURE 27: TYPES A, B AND C REGRESSIONS OF V ON R AND C IN THE V,C PLANE

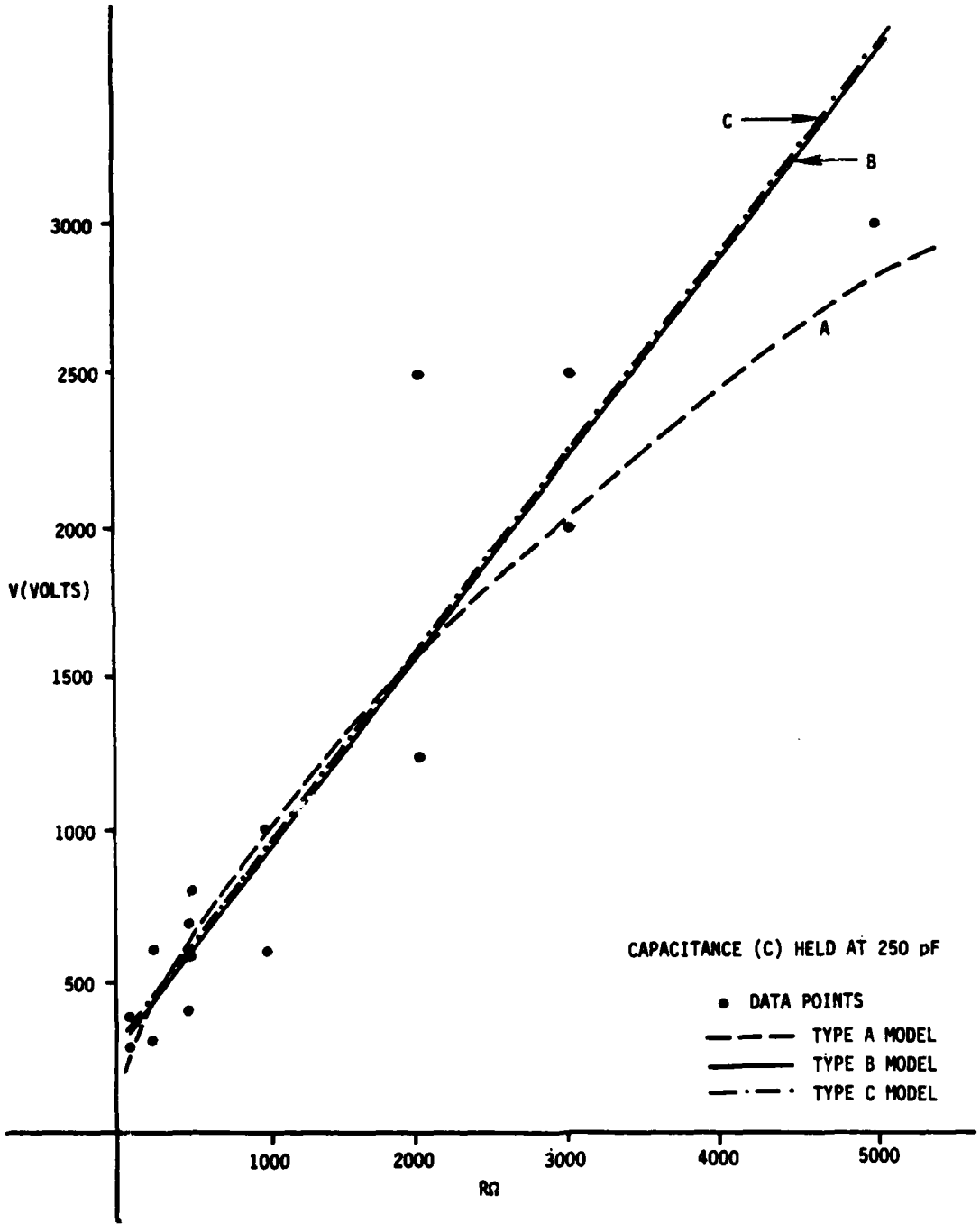



FIGURE 28: TYPES A, B, AND C REGRESSIONS OF V ON R AND C IN THE V,R PLANE

Unfortunately, in practical industrial and reliability work, data is rarely of the required structures. Least squares methods are applicable to departures from normality but the F statistic does require the normality assumption.

For data which is poorly measured and/or incorporates a high degree of noise on the measures, a non-parametric method may be preferable. The method due to Sen (Ref. 25) is suggested though its description and is outside the scope of this chapter. Recent work in the software analysis field at the Data and Analysis Center for Software (DACS) (Ref. 26) indicates that in certain instances the nonparametric method is preferable and goes on to show how symmetric and asymmetric noise affect the parametric and nonparametric regression methods.

## Qualitative (Discrete) Models

Variables like environments or package type (used when referring to component reliability) do not have numerical values as such and hence may be termed qualitative or discrete. When all the variables are discrete the analysis is defined in Chapter 2, although we have not yet solved the model numerically. We often require numerical solutions to problems involving both qualitative variables and continuous variables. Solutions are available by covariance analysis, described in the next chapter.

Particular attention should be paid to balanced data in constructing those models.

## Review

This note gives some basic information to enable the reader to carry out standard analyses and to understand the meaning of the terms in a computer printout from a regression analysis. Analysis of complex

data sets will probably require further methods and References 23, 24, 27 and 30 are recommended.

The extension of regression to solving analysis of variance type problems will be covered in Chapter 11.

# CHAPTER 10

## A Homogeneity Test for Noisy, Incomplete Data

CHAPTER 10

# A Homogeneity Test For Noisy, Incomplete Data

## Summary

This chapter gives a method and computer program for deciding whether data from diverse sources may be assumed to have come from the same distribution and therefore combined for analysis. Though there are standard tests in the literature, this note looks at a particular type of data which is prevalent in reliability studies. The method requires computer solution using a Monte Carlo simulation.

In some cases it may also be possible to determine the form of the underlying distribution but usually the censored structure of real reliability data will not allow such depth of analysis.

This note only considers the practical implementation; the theory is given in Ref. 33. The method is designed to encourage the analyst to look very critically at the data.

## Data

Suppose data is of the form

$r_1$ failures in time $t_1$
$r_2$ failures in time $t_2$
etc.
.
.
.
$r_m$ failures in time $t_m$

Then it is not possible to use a Weibull plot or standard analysis since individual failure times are not available.

Such data may arise either

(i) from a number of similar components operating simultaneously with $r_i$ failures discovered upon inspection after $t_i$ part hours (which is much greater than real time)

or

(ii) from a set of $r_i$ component times-to-failure in a total time of $t_i$, but individual failure times were lost for some administrative reason

Type (i) data will be termed "conventionally censored"; Type (ii) will be termed "artificially grouped."

Both conditions have arisen but this chapter is generally concerned only with conventionally censored data.

## Method

The principle of this method is to estimate the mean-time-to-failure (MTTF) from each record as $\hat{\theta}_i = t_i/r_i$. If there are m records then m estimates $\theta_1$, $\theta_2$ ... $\theta_m$ will result. These may then be plotted on a histogram thus

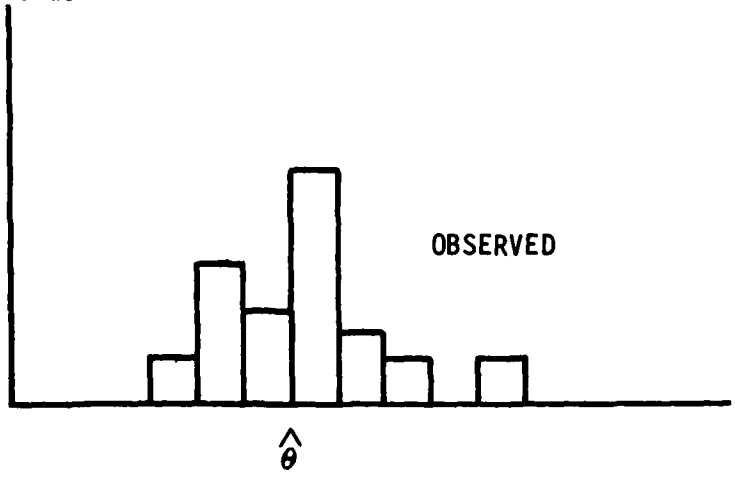



FIGURE 29: OBSERVED MTTFs

The theoretical distribution of $\hat{\theta}$ given that all the data are from the same underlying TTF distribution may also be found, which is one of the things the program does. (The mathematical derivation is given in the Appendix to this chapter). The theoretical distribution's typical shape is sketched below:

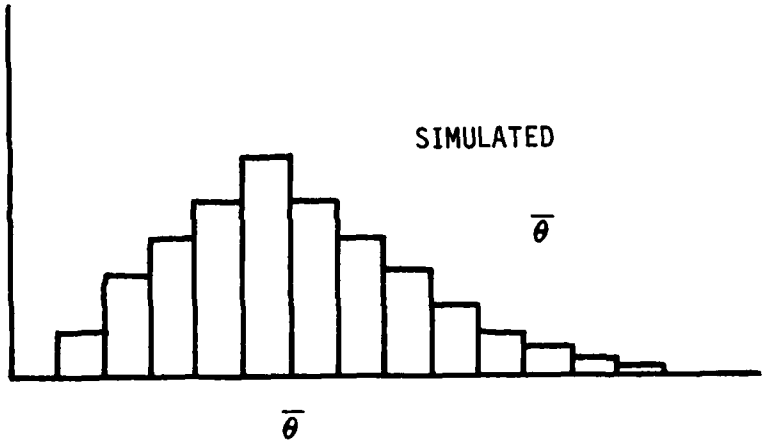



FIGURE 30: SIMULATED MTTFs

If the simulated distribution deviates significantly from the observed distribution (of $\hat{\theta}$) some non-homogeneity is indicated and the data may not be assumed to have all come from the same source.

Finally, a plot of $\hat{\theta}$ against r should exhibit a pronounced damping effect as r increases due to the central limit theorem as shown in Figure 29. If it does not, there may be a problem: one or more records could be outliers. Confidence intervals on each $\hat{\theta}$ should include $\bar{\theta}$ a given percentage $(1 - \alpha)\%$ of the time. This provides a further useful check on individual points and the data as a whole.

Data Analysis for Conventionally Censored Data

In such data, it is extremely difficult to cope with any underlying distribution other than exponential. However, for data from other distributions, the heavy censoring might intuitively be expected to produce a random process and hence the exponential TTF distribution would give a crude test on homogeneity.

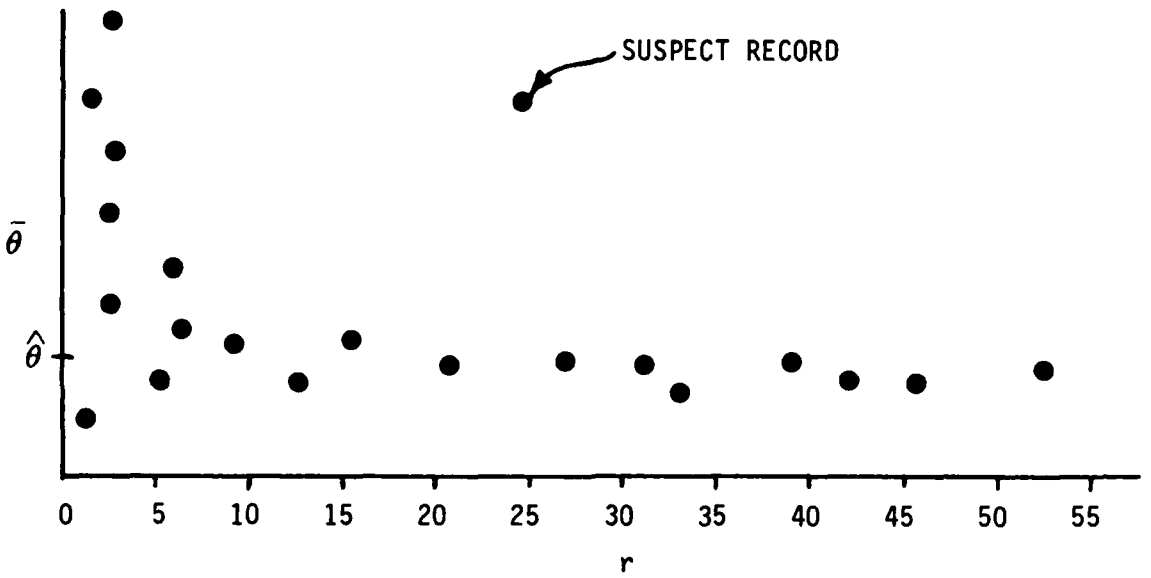



FIGURE 31: BEHAVIOR OF $\hat{\theta}$ WITH r

Distributions other than exponential are difficult to fit because the MTTF estimator for censored data is no longer straightforward.

The procedure is then as follows, using the program listed in the Appendix to this chapter:

> The program will ask for the number of records and a random number. Put these in first, separated by a space, then put in the data, one record per line as number of failures and number of part hours, separated by a space. Alternatively, the data may be read from a (similarly formatted) data file. Zero failure data may also be input. After all the records are input, the computer will suggest a histogram class interval; choose a convenient value near what it tells you. Incidentally, the random number generator is compatible with most machines.
>
> The program then outputs simulated and actual histograms of mean-times-to-failure evaluated from each record. % and cumulative % are also given. The % column may be used to

compare simulated to actual. If they are close then the data may be taken as "reasonably" homogeneous. The Kolmogorov-Smirnov statistic is then printed out with a 5% critical value to give an objective evaluation as to what is "reasonable." For small samples ($< 25$) this critical value is approximate; check in tables. $\hat{\theta}$ is also output - the overall MTTF estimate.

To consider the data more carefully, the program then outputs each r followed by the MTTF estimate for that record together with 60% confidence limits on the estimate. A more stringent test on the data results from the fact that if the data are homogeneous then 60% of the confidence intervals should include $\bar{\theta}$, which was output earlier.

Also, if Kolmogorov-Smirnov rejected the data as non-homogeneous, then the outliers would be suggested by those records whose confidence intervals most blatantly excluded $\bar{\theta}$.

If any data are rejected, then one or more records are not from the same underlying distribution or the data are not from an underlying exponential distribution.

Graphical displays of the results are also given. These routines may be handled on any terminal; the random number generator is portable.

Artificially Grouped Data

If, instead of being conventionally censored, the data are from actual (uncensored) component failure times but individual failure times are lost for administrative reasons then it may be possible to apply the method to other underlying distributions by modifying the program to generate samples from a nonexponential distribution. In practice, we

have found little use for such a refinement; users have generally found the program given adequate for their needs and/or data restrictions.

## Computer Outputs

Two annotated examples for conventionally censored data are now given:

```
$ RUN
$_File:       STAT10   EXAMPLE OF A SET OF HYDRAULIC VALVE DATA REJECTED BY THIS TEST
IS THE INPUT DATA TO STAT10 CONTAINED IN A FILE?  Y / N

Y
ENTER THE NAME OF THE DATA FILE

ST10.DAT
NUMBER OF INPUT RECS IS         35    THE RANDOM NUMBER IS      10001
       0          0.2120
       0          0.0530
      14          0.0530
       9          0.2120
      37          0.2120
       1          0.1060
       0          0.2650
      64          1.3250
       0          0.0530
      14          0.0580
       0          0.5300
       4          0.0530
      48          1.7400
      10          0.8700
       8          0.4350
       7          0.4350
      54          1.3050
     238          1.7400
       0          0.1340
     109          6.4324
       1          0.0500
       1          0.4350
       6          1.8700
      18          0.8700
       2          0.4350
       3          0.4350
       3          1.4350
      10          0.4350
     108          0.8700
      67          1.3050
       0          0.4350
       2          0.4350
       8          0.8700
       3          1.3050
       5          0.4350
```

*The data is input in pairs:*
*No. Failures; No. Part Hours.*

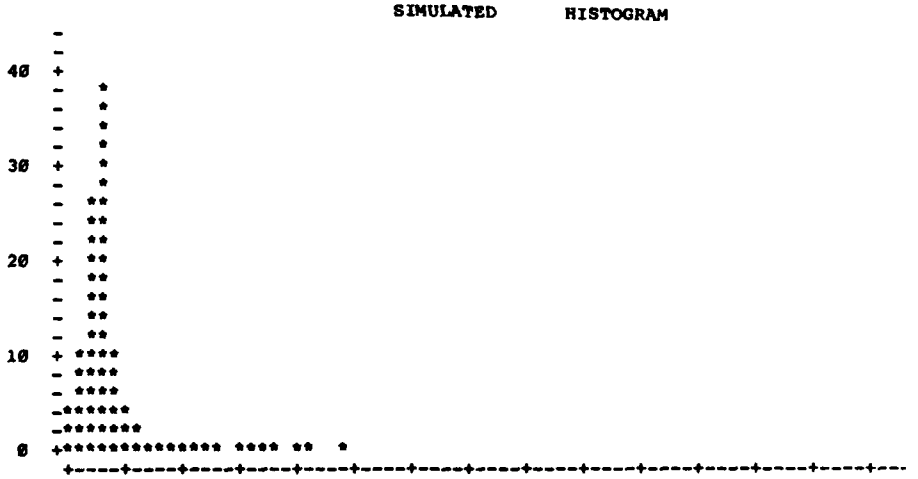
```
HISTOGRAM CLASS INTERVAL? SUGGEST ABOUT  0.01

0.01

STAT10 IS PROCESSING
```

| | | SIMULATED | | | ACTUAL | | |
|---|---|---|---|---|---|---|---|
| CLASS INTERVAL | | FREQ. | % | CUM% | FREQ. | % | CUM% |
| 0.00- | 0.01 | 428.0 | 5.10 | 5.10 | 5.00 | 17.86 | 17.86 |
| 0.01- | 0.02 | 896.0 | 10.67 | 15.76 | 2.00 | 7.14 | 25.00 |
| 0.02- | 0.03 | 2203.0 | 26.23 | 41.99 | 3.00 | 10.71 | 35.71 |
| 0.03- | 0.04 | 3225.0 | 38.39 | 80.38 | 1.00 | 3.57 | 39.29 |
| 0.04- | 0.05 | 893.0 | 10.63 | 91.01 | 2.00 | 7.14 | 46.43 |
| 0.05- | 0.06 | 346.0 | 4.12 | 95.13 | 3.00 | 10.71 | 57.14 |
| 0.06- | 0.07 | 181.0 | 2.15 | 97.29 | 1.00 | 3.57 | 60.71 |
| 0.07- | 0.08 | 87.0 | 1.04 | 98.32 | 0.00 | 0.00 | 60.71 |
| 0.08- | 0.09 | 38.0 | 0.45 | 98.77 | 2.00 | 7.14 | 67.86 |
| 0.09- | 0.10 | 42.0 | 0.50 | 99.27 | 0.00 | 0.00 | 67.86 |
| 0.10- | 0.11 | 20.0 | 0.24 | 99.51 | 2.00 | 7.14 | 75.00 |
| 0.11- | 0.12 | 15.0 | 0.18 | 99.69 | 0.00 | 0.00 | 75.00 |
| 0.12- | 0.13 | 10.0 | 0.12 | 99.81 | 0.00 | 0.00 | 75.00 |
| 0.13- | 0.14 | 3.0 | 0.04 | 99.85 | 0.00 | 0.00 | 75.00 |
| 0.14- | 0.15 | 0.0 | 0.00 | 99.85 | 1.00 | 3.57 | 78.57 |
| 0.15- | 0.16 | 3.0 | 0.04 | 99.88 | 0.00 | 0.00 | 78.57 |
| 0.16- | 0.17 | 2.0 | 0.02 | 99.90 | 0.00 | 0.00 | 78.57 |
| 0.17- | 0.18 | 2.0 | 0.02 | 99.93 | 0.00 | 0.00 | 78.57 |
| 0.18- | 0.19 | 1.0 | 0.01 | 99.94 | 0.00 | 0.00 | 78.57 |
| 0.20- | 0.21 | 2.0 | 0.02 | 99.96 | 0.00 | 0.00 | 78.57 |
| 0.21- | 0.22 | 2.0 | 0.02 | 99.99 | 2.00 | 7.14 | 85.71 |
| 0.24- | 0.25 | 1.0 | 0.01 | 100.00 | 0.00 | 0.00 | 85.71 |
| 0.31- | 0.32 | 0.0 | 0.00 | 100.00 | 1.00 | 3.57 | 89.29 |
| 0.43- | 0.44 | 0.0 | 0.00 | 100.00 | 2.00 | 7.14 | 96.43 |
| 0.47- | 0.48 | 0.0 | 0.00 | 100.00 | 1.00 | 3.57 | 100.00 |

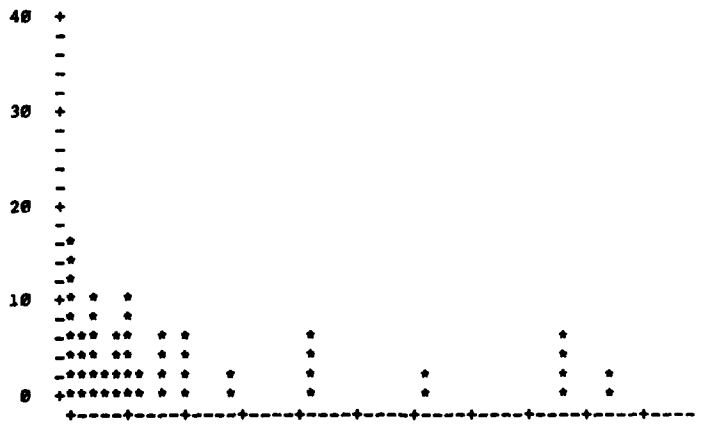
DO YOU WISH HISTOGRAMS OF THE SIMULATED AND ACTUAL DATA?
Y

*THESE HISTOGRAMS GIVE A QUICK VISUAL APPRECIATION.*
*AXES ARE FREQUENCY VS. TIME. THE ABCISSA SCALE IS*
*ARBITRARILY LABELED.*

**SIMULATED HISTOGRAM**




**ACTUAL HISTOGRAM**



THETA= 0.0326 *(THIS IS THE MTTF ESTIMATE OVERALL)*
*KOLMOGOROV-SMIRNOV TEST FOLLOWS:*

DMAX= 0.440. CRITICAL VALUE AT 5% SIGNIFICANCE 0.257

*(SO THE DATA ARE REJECTED AS BEING FROM THE SAME SOURCE AS DMAX >0.257)*

N.B. CRITICAL VALUE IS NOT TOO ACCURATE FOR SMALL SAMPLE SIZES.
CHECK TABLES FOR SAMPLE SIZES < 30.

**********

BEHAVIOR OF MTTF ESTIMATE WITH # FAILURES

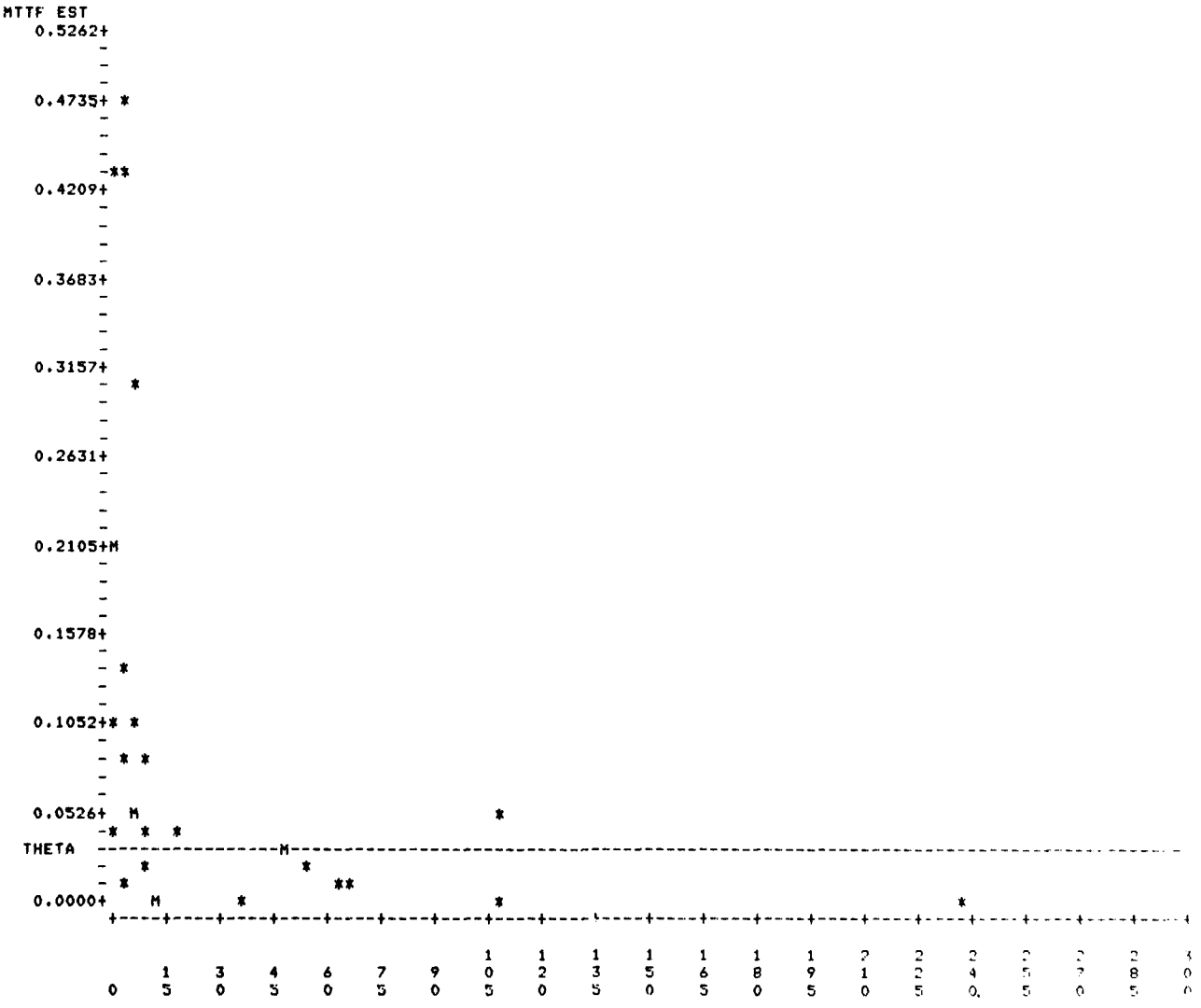
| NO. FAILURES | MTTF EST. | *CONFIDENCE INTERVALS* LL | UL |
|---|---|---|---|
| 14.0 | 0.0038 | 0.0029 | 0.0049 |
| 9.0 | 0.0236 | 0.0169 | 0.0329 |
| 37.0 | 0.0057 | 0.0049 | 0.0067 |
| 1.0 | 0.1060 | 0.0349 | 0.5350 |
| 64.0 | 0.0207 | 0.0185 | 0.0232 |
| 14.0 | 0.0041 | 0.0032 | 0.0054 |
| 4.0 | 0.0132 | 0.0078 | 0.0231 |
| 48.0 | 0.0362 | 0.0317 | 0.0413 |
| 10.0 | 0.0870 | 0.0636 | 0.1192 |
| 8.0 | 0.0544 | 0.0381 | 0.0779 |
| 7.0 | 0.0621 | 0.0424 | 0.0918 |
| 54.0 | 0.0242 | 0.0213 | 0.0273 |
| 238.0 | 0.0073 | 0.0069 | 0.0077 |
| 109.0 | 0.0590 | 0.0542 | 0.0642 |
| 1.0 | 0.0500 | 0.0164 | 0.2523 |
| 1.0 | 0.4350 | 0.1431 | 2.1954 |
| 6.0 | 0.3117 | 0.2052 | 0.4784 |
| 18.0 | 0.0483 | 0.0385 | 0.0605 |
| 2.0 | 0.2175 | 0.1006 | 0.5347 |
| 3.0 | 0.1450 | 0.0783 | 0.2841 |
| 3.0 | 0.4783 | 0.2582 | 0.9372 |
| 10.0 | 0.0435 | 0.0318 | 0.0596 |
| 108.0 | 0.0081 | 0.0074 | 0.0088 |
| 67.0 | 0.0195 | 0.0174 | 0.0217 |
| 2.0 | 0.2175 | 0.1006 | 0.5347 |
| 8.0 | 0.1088 | 0.0762 | 0.1558 |
| 3.0 | 0.4350 | 0.2348 | 0.8523 |
| 5.0 | 0.0870 | 0.0547 | 0.1407 |

*Listing of Statistics for each record.*

DO YOU WISH A PLOT OF MTTF VS. FAILURES?

YES

PLOT OF BEHAVIOUR OF MTTF EXTIMATE VERSUS THE NUMBER OF FAILURES




HORIZONTAL SCALE 3
VERTICAL SCALE 0.0105
THETA 0.0326

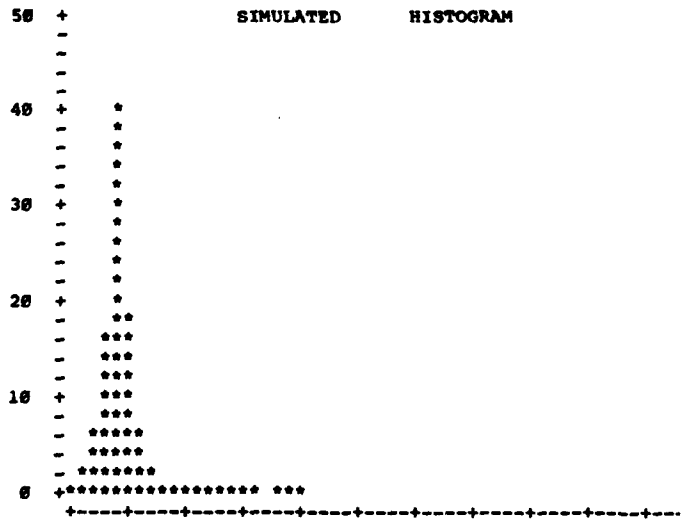
FORTRAN STOP
$

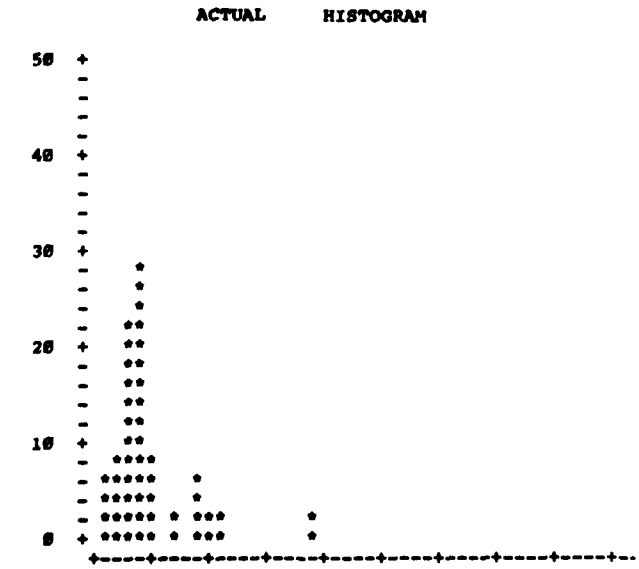
*EXAMPLE OF A SET OF TTL DATA, ACCEPTED*

```
  RUN
$ File:       STAT10
IS THE INPUT DATA TO STAT10 CONTAINED IN A FILE?  Y / N

N
INPUT THE # OF DATA RECORDS FOLLOWED BY A RANDOM NUMBER

30 26758
ENTER THE DATA, TWO ENTRIES PER LINE        Manual Input

19 80.098
9 92.716
24 86.59
40 123.301
15 53.468
7 53.173
43 176.415
55 145.297
5 97.144
142 485.628
2 2.617
3 28.141
6 24.300
2 22.329
11 62.868
84 300.909
142 783.401
5 48.754
31 139.714
40 135.291
122 565.891
323 1855.787
9 10.175
32 137.68
13 60.819
24 87.464
50 221.918
28 137.729
25 71.996
5 14.835

HISTOGRAM CLASS INTERVAL? SUGGEST ABOUT  1.39

1.00

STAT10 IS PROCESSING
```

| | SIMULATED | | | ACTUAL | | |
|---|---|---|---|---|---|---|
| CLASS INTERVAL | FREQ. | % | CUM% | FREQ. | % | CUM% |
| 0.00- 1.00 | 64.0 | 0.71 | 0.71 | 0.00 | 0.00 | 0.00 |
| 1.00- 2.00 | 209.0 | 2.32 | 3.03 | 2.00 | 6.67 | 6.67 |
| 2.00- 3.00 | 571.0 | 6.34 | 9.38 | 3.00 | 10.00 | 16.67 |
| 3.00- 4.00 | 1610.0 | 17.89 | 27.27 | 7.00 | 23.33 | 40.00 |
| 4.00- 5.00 | 3752.0 | 41.69 | 68.96 | 9.00 | 30.00 | 70.00 |
| 5.00- 6.00 | 1768.0 | 19.64 | 88.60 | 3.00 | 10.00 | 80.00 |
| 6.00- 7.00 | 552.0 | 6.13 | 94.73 | 0.00 | 0.00 | 80.00 |
| 7.00- 8.00 | 228.0 | 2.53 | 97.27 | 1.00 | 3.33 | 83.33 |
| 8.00- 9.00 | 108.0 | 1.20 | 98.47 | 0.00 | 0.00 | 83.33 |
| 9.00- 10.00 | 65.0 | 0.72 | 99.19 | 2.00 | 6.67 | 90.00 |
| 10.00- 11.00 | 24.0 | 0.27 | 99.46 | 1.00 | 3.33 | 93.33 |
| 11.00- 12.00 | 17.0 | 0.19 | 99.64 | 1.00 | 3.33 | 96.67 |
| 12.00- 13.00 | 13.0 | 0.14 | 99.79 | 0.00 | 0.00 | 96.67 |
| 13.00- 14.00 | 5.0 | 0.06 | 99.84 | 0.00 | 0.00 | 96.67 |
| 14.00- 15.00 | 6.0 | 0.07 | 99.91 | 0.00 | 0.00 | 96.67 |
| 15.00- 16.00 | 3.0 | 0.03 | 99.94 | 0.00 | 0.00 | 96.67 |
| 16.00- 17.00 | 1.0 | 0.01 | 99.96 | 0.00 | 0.00 | 96.67 |
| 18.00- 19.00 | 1.0 | 0.01 | 99.97 | 0.00 | 0.00 | 96.67 |
| 19.00- 20.00 | 1.0 | 0.01 | 99.98 | 1.00 | 3.33 | 100.00 |
| 20.00- 21.00 | 2.0 | 0.02 | 100.00 | 0.00 | 0.00 | 100.00 |

DO YOU WISH HISTOGRAMS OF THE SIMULATED AND ACTUAL DATA?
YES



THETA= 4.6402

DMAX= 0.150. CRITICAL VALUE AT 5% SIGNIFICANCE 0.248

*(Data is accepted since DMAX is <0.248)*

N.B. CRITICAL VALUE IS NOT TOO ACCURATE FOR SMALL SAMPLE SIZES.
CHECK TABLES FOR SAMPLE SIZES < 30.

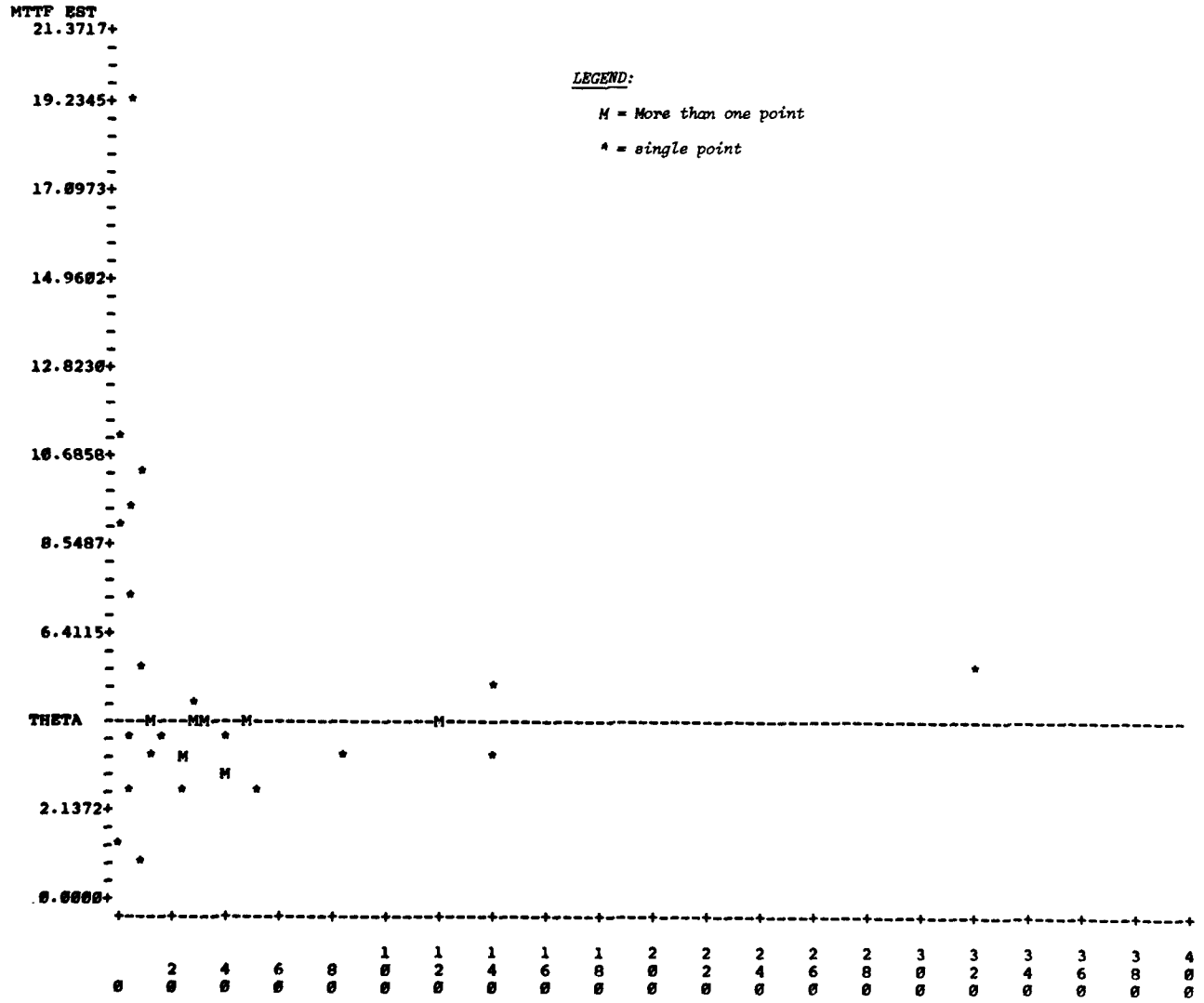
**********
BEHAVIOR OF MTTF ESTIMATE WITH # FAILURES

| NO.FAILURES | MTTF EST. | LL | UL |
|---|---|---|---|
| 19.0 | 4.2157 | 3.3842 | 5.2413 |
| 9.0 | 10.3018 | 7.3844 | 14.4034 |
| 24.0 | 3.6079 | 2.9741 | 4.3678 |
| 40.0 | 3.0825 | 2.6630 | 3.5616 |
| 15.0 | 3.5645 | 2.7750 | 4.5722 |
| 7.0 | 7.5961 | 5.1772 | 11.2177 |
| 43.0 | 4.1027 | 3.5643 | 4.7140 |
| 55.0 | 2.6418 | 2.3355 | 2.9836 |
| 5.0 | 19.4288 | 12.2260 | 31.4126 |
| 142.0 | 3.4199 | 3.1744 | 3.6817 |
| 2.0 | 1.3085 | 0.6054 | 3.2167 |
| 3.0 | 9.3803 | 5.0638 | 18.3784 |
| 6.0 | 4.0500 | 2.6662 | 6.2170 |
| 2.0 | 11.1645 | 5.1650 | 27.4457 |
| 11.0 | 5.7153 | 4.2442 | 7.6977 |
| 84.0 | 3.5822 | 3.2469 | 3.9477 |
| 142.0 | 5.5169 | 5.1208 | 5.9391 |
| 5.0 | 9.7508 | 6.1359 | 15.7652 |
| 31.0 | 4.5069 | 3.8100 | 5.3206 |
| 40.0 | 3.3823 | 2.9220 | 3.9079 |
| 122.0 | 4.6385 | 4.2788 | 5.0239 |
| 323.0 | 5.7455 | 5.4731 | 6.0291 |
| 9.0 | 1.1306 | 0.8104 | 1.5807 |
| 32.0 | 4.3025 | 3.6477 | 5.0648 |
| 13.0 | 4.6784 | 3.5674 | 6.1301 |
| 24.0 | 3.6443 | 3.0041 | 4.4119 |
| 50.0 | 4.4384 | 3.8985 | 5.0448 |
| 28.0 | 4.9189 | 4.1186 | 5.8626 |
| 25.0 | 2.8798 | 2.3841 | 3.4715 |
| 5.0 | 2.9670 | 1.8670 | 4.7971 |

DO YOU WISH A PLOT OF MTTF VS. FAILURES?

YES

PLOT OF BEHAVIOUR OF MTTF EXTIMATE VERSUS THE NUMBER OF FAILURES




HORIZONTAL SCALE 4
VERTICAL SCALE 0.4274
THETA 4.6402

FORTRAN STOP

## APPENDIX TO CHAPTER 10

This Appendix lists the program and subroutines used to run the simulation and evaluate the data.

STAT-10
MAIN PROGRAM

```
   T\
$ TYPE STAT10.FOR
C
      REAL DATA(100)
      REAL HISTO(100)
      REAL FAIL(100)
      REAL TIME(100)
      REAL THAT(100)
      REAL NDF
       CHARACTER * 1 ANS
       CHARACTER * 15 FILNM
       INTEGER SEED
C
C     STAT10 EVALUATES A SET OF FAILURE RECORDS TO SEE WHETHER THEY MAY
C     STATISTICALLY BE ASSUMED TO HAVE COME FROM THE SAME DISTRIBUTION.
C     OUTPUT IS A TEST ON EACH RECORD TOGETHER WITH AN OVERALL
C     GOODNESS OF FIT TEST ON THE COMPLETE DATA SET.THE PROGRAM IS
C     DESIGNED TO ENCOURAGE THE ANALYST TO LOOK CAREFULLY AT THE DATA
C     AND UNDERSTAND IT. SEE STAT NOTE #10 FOR DETAILS AND
C     INSTRUCTIONS.
C       FOR A MORE COMPLETE DISCUSSION, SEE "DEY,K.A. 'STATISTICAL
C       ANALYSIS OF NOISY, INCOMPLETE FAILURE DATA', PROCEEDINGS
C       R AND M SYMPOSIUM 1982.
C
C
C       A - RANDOM NUMBER RETURNED BY THE FUNCTION RANNOS
C       AINT - CONVERT A REAL TO AN INTEGER (ABSOLUTE VALUE)
C       APPX - RESULT OF APPLYING THE APPROXIMATION OF THE
C              CHI SQUARE FORMULA
C       ANS - VALUE ENTERED BY THE USER FROM THE TERMINAL
C             = YES, USER DAT IS SUPPLIED FROM AN EXTERNAL FILE
C             = NO, USER ENTERS DATA DURING PROGRAM EXECUTION
C       CHIL - THE LOWER CONFIDENCE LIMIT OF THE DATA GENERATED
C              BY THE CHI SQUARE FORMULA
C       CHIU - THE UPPER CONFIDENCE LIMIT OF THE DATA GENERATED
C              BY THE CHI SQUARE FORMULA
C       CUM - CUMULATIVE TOTAL (PERCENT) OF THE SIMULATED
C             HISTOGRAM DATA (I.E. THE C.D.F.)
C       DATA - HISTOGRAM ARRAY FOR THE ACTUAL DATA
C       DELTA - APPROXIMATE CLASS INTERVAL SIZE FOR A HISTOGRAM
C               GENERATED FROM THE INPUT DATA; USER CAN MODIFY
C               THIS VALUE .
C       DMAX - THE MAXIMUM DEVIATION OF THE USER DATA (C.D.F) FROM THE
C              SIMULATED C.D.F.
C       FAIL - AN ARRAY CONTAINING THE NUMBER OF FAILURES PER TIME(T)
C              DATA IS IN THE ORDER INPUT TO STAT10
C       FAILS - TOTAL NUMBER OF FAILURES INPUT
C       FILNM - NAME OF THE FILE CONTAINING THE USER INPUT DATA
C       HISTO - HISTOGRAM ARRAY FOR THE SIMULATED DATA
C       ICLASS - HISTOGRAM CLASS INTERVAL (INDEX TO THE BUCKETS)
C                FOR SIMULATED DATA
C       IFILE - FILE CODE OF THE USER INPUT FILE; =3, DATA IS TO BE
C               READ FROM A FILE; =5, DATA IS TO BE READ FROM A
C               TERMINAL
C       IHAT - HISTOGRAM CLASS INTERVAL (IE INDEX TO THE BUCKETS)
C              FOR THE ACTUAL DATA
C       ISSZ - CONVENIENT SAMPLE SIZE FOR THE SIMULATION (=300)
C       ITOT - TOTAL NUMBER OF SAMPLES GENERATED, = 300*NREC
C       JR - NUMBER OF FAILURES FOR A TIME (T)
C       JRR - USER INPUT OF NUMBER OF FAILURES PER TIME (T)
C       KS - RESULT OF THE GOODNESS OF FIT CALCULATION
C       NDF - PARAMETER FOR CRITICAL K-S STATISTICAL VALUES
C       NREC - NUMBER OF USER INPUT DATA RECORDS OR ENTRIES
C       NZ - NUMBER OF USER ENTRIES WHICH HAD ZERO FAILURES IN
C            IN THE OBSERVED TIME
C       PCENT - INTERVAL PERCENTILE FOR SIMULATION
C       PHRS - TOTAL NUMBER OF FAILURE HOURS
C       QCENT - INTERVAL PERCENTILE FOR DATA
C       QCUM - CUMULATIVE TOTAL OF PERCENT OF THE ACTUAL HISTO DATA
C       RANNOS - FUNCTION CALLED TO GENERATE A RANDOM NUMBER
C       SEED - SEED VALUE FOR THE RANDOM NUMBER GENERATOR RANNOS;
C              VALUE IS INTEGER BETWEEN 1 AND 999,999,999
C       T - TIME INTERVAL FOR THE NUMBER OF USER OBSERVED FAILURES
C       THAT - AN ARRAY CONTAINING THE RATIO OF JRR/T (MEAN TIME TO
C              FAILURE-MTTF) FOR EACH USER INPUT OF JRR WHICH IS
C              NON-ZERO AND T. ORDER OF THAT IS THE SAME AS THE
C              USER ORDER OF INPUT FOR JRR AND T.
C       THETA - AVERAGE OF TOTAL HOURS / TOTAL FAILURES
C       TIME - AN ARRAY CONTAINING THE TIME INTERVALS FOR USER
C              OBSERVED FAILURES.  (ORDER OF TIME IS THE SAME AS
C              THE USER ORDER OF INPUT FOR JRR AND T.)
C       UPPER - UPPER VALUE OF A CLASS INTERVAL OF A HISTOGRAM
C       X - COMPUTED EXPONENTIAL VALUE
C       XLOW - LOWER VALUE FO A CLASS INTERVAL OF A HISTOGRAM
C       Y - THE MEAN OF THE JR EXPONENTIAL DISTRIBUTION
C       INITIALIZE CONSTANTS
C
      DMAX=0
      FAILS=0
      PHRS=0
```

```
C
C       WRITE A MESSAGE TO THE USER ASKING IF HE WISHES TO INPUT HIS
C       DATA DIRECTLY OR IF THE DATA IS CONTAINED IN A FILE
C       FOR USE BY STAT10
C
        WRITE (6,1000)
C
C       READ HIS REPLY
C
        READ (5,1020) ANS
C
C       IS THE DATA ON A FILE
C
        IF (ANS .EQ. 'Y') GO TO 2
C
C       NO, USER WISHES TO ENTER THE DATA DIRECTLY,PROMPT HIM
C       AND READ THE DATA.  THE FIRST INPUTS ARE THE # OF
C       DATA ENTRIES AND A RANDOM NUMBER
C
        WRITE (6,1100)
C
C       READ THE NUMBER OF DATA RECORDS (NREC) AND THE RANDOM
C       NUMBER (RANNUM)
C
        READ (5,*) NREC, SEED
C
C       PROMPT THE USER TO ENTER THE DATA
C
        WRITE (6,1150)
C
C       READ THE USER INPUT FROM THE TERMINAL - IE IFILE = 5
C
        IFILE = 5
        GO TO 3
C
C       ASK THE USER TO ENTER THE NAME OF THE FILE ON WHICH
C       HIS DATA RESIDES
C
  2     WRITE (6,1300)
C
C       READ HIS REPLY
C
  21    READ (5,1350) FILNM
C
C       OPEN THE FILE AND READ THE DATA - PROVIDE AND ERROR
C       RETURN TO THE USER IN CASE HE MISSTYPES - #400
C
  22    OPEN (UNIT=3,NAME=FILNM,STATUS='OLD',ERR=400)
C
C       GET THE NUMBER OF RECORDS AND THE RANDOM NUMBER
C
        READ (3,*,END=400,ERR=400)NREC , SEED
C
C       READ THE DATA RECORDS AND INPUT PROCESS
C       PRINT THE NUM OV RECS AND THE RAN NUM FOR THE USER
C
        WRITE (6,1360)NREC,SEED
C
C       SET THE INPUT FILE TO THE DISK FILE
C
        IFILE = 3
C
C
C
C
C       PROCESS THE INPUT DATA
C
C
C
  3     CONTINUE
      DO 4 IREK=1,NREC
      READ(IFILE,*) JRR,T
C
C       IF THE INPUT IS CONTAINED ON A DATA FILE, PRINT
C       IT AS IT IS READ
C
        IF (IFILE .EQ. 3) WRITE (6,1375)JRR,T
      THAT(IREK)=0
      FAIL(IREK)=JRR
      TIME(IREK)=T
      IF(JRR.LT.1) GO TO 4
      THAT(IREK)=(T/JRR)
C
C       IF THE INPUT DATA IS BEING READ FROM A FILE, NO USER
C       PROMPT IS NEEDED
        IF (IFILE .EQ. 3) GO TO 4
C
C       THE INPUT DATA IS BEING READ FROM A TERMINAL, PROMPT
C       THE USER FOR THE NEXT DATA LINE.
C         REMEMBER NOT TO PROMPT AFTER THE LAST LINE IS INPUT
C
```

```
        IF (IREC .EQ. NREC) GO TO 4
C       WRITE (6,1200)
    4 CONTINUE
      QCUM=0
      NZ=0
      ITOT=0
      CUM=0
C
C       INITIALIZE THE HISTO AND DATA ARRAYS
C
      DO 5 JCLR=1,100
      HISTO(JCLR)=0
      DATA(JCLR)=0
    5 CONTINUE
C
C       COMPUTE THE TOTAL # OF FAILURES AND THE TOTAL # OF HOURS
C
      DO 10 IMLE =1,NREC
      FAILS=FAILS+FAIL(IMLE)
      PHRS=PHRS+TIME(IMLE)
   10 CONTINUE
C
C       COMPUTE THE AVERAGE TIME BETWEEN FAILURES (THETA)
C       COMPUTE THE BUCKET SIZE (INTERVAL) FOR THE HISTOGRAM AND
C         SUGGEST IT TO THE USER.  READ HIS REPLY AND USE HIS VALUE
C
      THETA=PHRS/FAILS
      DELTA=.299*THETA
      WRITE(6,11) DELTA
      READ(5,*) DELTA
C
C       TELL THE USER THAT STAT10 IS PROCESSING
C
        WRITE (6,1400)
C
C       INITIALIZE CONVENIENT SAMPLE SIZE TO 300
C
      ISSZ=300
      DO 140 ILOP=1,NREC
C
C       DETERMINE THE # OF USER INPUTS IN WHICH NO FAILURES
C       WERE OBSERVED (NZ)
C
      JR=FAIL(ILOP)
      IF(JR.GT.0) GO TO 14
      NZ=NZ+1
      GO TO 140
C
C       GENERATE A RANDOM SAMPLE FOR AN EXPONENTIAL DISTRIBUTION
C       ONE POINT AT A TIME FOR EACH NON-ZERO # OF FAILURE VALUE
C
   14 DO 100 ISAMP=1,ISSZ
      Y=0
      DO 90 IR=1,JR
C
C       COMPUTE A RANDOM NUMBER
C
      A=RANNOS(SEED)
C
C       COMPUTE THE EXPONENTIAL
C
      X=-(THETA*ALOG(1-A))
      Y=Y+X
   90 CONTINUE
C
C       DETERMINE THE MEAN OF THIS FAILURE EXPONENTIAL
C
      Y=Y/JR
C
C       COMPUTE AN INDEX TO DETERMINE WHICH BUCKET OF THE HISTOGRAM
C       FOR THE GENERATED DATA SHOULD BE INCREMENTED AND
C       INCREMENT IT
C
      ICLASS=AINT(Y/DELTA)+1
      HISTO(ICLASS)=HISTO(ICLASS)+1
  100 CONTINUE
C
C       KEEP A RUNNING COUNT OF THE TOTAL # OF SAMPLES GENERATED
C       COMPUTE AN INDEX TO DETERMINE WHICH BUCKET OF THE HISTOGRAM
C       FOR THE USER INPUT DATA SHOULD BE INCREMENTED AND
C       INCREMENT IT.
C
      ITOT=ITOT+ISSZ
      IHAT=AINT(THAT(ILOP)/DELTA)+1
      DATA(IHAT)=DATA(IHAT)+1
  140 CONTINUE
```

```
C
C       WRITE THE HISTOGRAM HEADERS
C
  150 WRITE(6,161)
      WRITE(6,162)
C
C       PRINT THE HISTOGRAM DATA;  IE THE CLASS INTERVAL (BUCKET)
C       AND ITS FREQUENCY, PERCENT, AND CUMULATIVE PERCENT FOR THE
C       SIMULATED AND ACTUAL DATA.
C
      XLOW=0
      DO 205 IH=1,100
      UPPER=XLOW+DELTA
      IF(HISTO(IH).EQ.0.AND.DATA(IH).EQ.0) GO TO 204
C
      PCENT=(100*HISTO(IH))/ITOT
      CUM=CUM+PCENT
C
      QCENT=(100*DATA(IH))/(NREC-NZ)
      QCUM=QCUM+QCENT
C       PRINT THE LINE
      WRITE(6,203) XLOW,UPPER,HISTO(IH),PCENT,CUM,DATA(IH),QCENT,QCUM
      KS=ABS(CUM-QCUM)
      IF(KS.LT.DMAX) GO TO 204
      DMAX=KS
  204 XLOW=UPPER
  205 CONTINUE
C
C       ASK THE USER IF THEY WISH TO PLOT THE ACTUAL AND SIMULATED
C       DATA IN A HISTOGRAM
C
        WRITE (06,2100)
C
C       READ THE REPLY
C
        READ (05,1020) ANS
C
        IF (ANS .EQ. 'N') GO TO 275
C
C       YES PLOT THE HISTOGRAM FOR SIMULATED AND THEN THE
C       HISTOGRAM FOR ACTUAL DATA.
C
        CALL HSTGRM (HISTO,DELTA,'SIMULATED     ')
C
        CALL HSTGRM (DATA,DELTA,'    ACTUAL     ')
C
  275   CONTINUE
C
C       PRINT THETA,MAX DEVIATION, AND CRITICAL DATA
C
      WRITE(6,207) THETA
      NDF=ITOT/300
      CRIT=1.36/(SQRT(NDF))
      DMAX=DMAX/100
      WRITE(6,208) DMAX,CRIT
C
C       OUTPUT A WARNING MESSAGE FOR SMALL SAMPLE SIZES.
C
        WRITE (06,209)
      WRITE(6,210)
C       THE FOLLOWING TABLE GIVES THE NUMBER OF FAILURES PER RECORD
C       (R), THE CORRESPONDING MTTF ESTIMATE TOGETHER WITH THE LOWER
C       AND UPPER 60% CONFIDENCE LIMITS (LL AND UL) ASSUMING
C       EXPONENTIAL DATA.
C       THE MTTF ESTIMATE IS THE PLOTTED AGAINST 'R', WITH A DOTTED
C       LINE THROUGH THE PLOT AT THETA, WHERE THETA IS THE OVERALL
C       MAXIMUM LIKELIHOOD ESTIMATE FOR ALL THE DATA.
C
      WRITE(6,250)
      WRITE(6,260)
      DO 300 IPLOT=1,NREC
      IF(FAIL(IPLOT).LT.1) GO TO 300
      APPX=(0.84179+(SQRT(3+(4*FAIL(IPLOT)))))**2
      CHIL=(4*TIME(IPLOT))/APPX
      APPX=(-0.84179+(SQRT((4*FAIL(IPLOT))-1)))**2
      CHIU=(4*TIME(IPLOT))/APPX
      WRITE(6,270) FAIL(IPLOT),THAT(IPLOT),CHIL,CHIU
  300 CONTINUE
```

```
C
C       ASK THE USER IF THEY WISH A PLOT OF FAILURES VS MTTF
C
        WRITE (06,2200)
C
C       READ THEIR REPLY
C
        READ (05,1020)ANS
C
        IF (ANS .EQ. 'N') GO TO 350
C
C       CALL TPLOT TO CREATE THE PLOT
C

        CALL TPLOT (FAIL,THAT,THETA)
 350    CONTINUE
      STOP
C
C       USER HAS MIS-TYPED THE NAME OF THE DATA FILE
C       TELL HIM AND GIVE HIM ANOTHER CHANCE
C
  400   WRITE (6,2000)
        GO TO 21
C
C
C       FORMAT STATEMENTS
C
C
   11 FORMAT(1X,//,' HISTOGRAM CLASS INTERVAL? SUGGEST ABOUT',F6.2,//)
  161  FORMAT (24X,'SIMULATED',15X,'ACTUAL')
  162 FORMAT(3X,'CLASS INTERVAL      FREQ.     %     CUM%
     &     FREQ.    %      CUM%')
  203 FORMAT(1X,F6.2,'-',F6.2,'      ',F8.1,2F8.2,3F8.2)
  207 FORMAT(1X,//,' THETA=',F8.4,//)
  208 FORMAT(1X,'DMAX=',F6.3.'. CRITICAL VALUE AT 5% SIGNIFICANCE
     & ',F5.3,//)
 209    FORMAT (//,' N.B. CRITICAL VALUE IS NOT TOO ACCURATE',
     1  ' FOR SMALL SAMPLE SIZES.',/,' CHECK TABLES FOR SAMPLE',
     2  ' SIZES < 30.',//)
  210 FORMAT(1X,'************************************************')
  250 FORMAT(1X,'BEHAVIOR OF MTTF ESTIMATE WITH # FAILURES',1X,//)
  260 FORMAT(1X,'NO.FAILURES   MTTF EST.      LL
     &          UL       ',//)
  270 FORMAT(1X,F5.1,'          ',F8.4,'    ',2F8.4)
 1000   FORMAT(' IS THE INPUT DATA TO STAT10 CONTAINED IN',
     1 ' A FILE?  Y / N',//)
 1020   FORMAT (A1)
 1100   FORMAT (' INPUT THE # OF DATA RECORDS FOLLOWED BY A',
     1  ' RANDOM NUMBER',//)
 1150   FORMAT (' ENTER THE DATA, TWO ENTRIES PER LINE',/)
 1300   FORMAT (' ENTER THE NAME OF THE DATA FILE',//)
 1350   FORMAT (A15)
 1360   FORMAT (' NUMBER OF INPUT RECS IS  ',I8,
     1  '    THE RANDOM NUMBER IS  ',I8)
 1375   FORMAT (I8,3X,F12.4)
 1400   FORMAT (1X,//,' STAT10 IS PROCESSING',//)
 2000   FORMAT (' FILE NAME IS INCORRECT, PLEASE RE-ENTER',//)
 2100   FORMAT (///,' DO YOU WISH HISTOGRAMS OF THE SIMULATED',
     1  ' AND ACTUAL DATA?')
 2200   FORMAT (//,' DO YOU WISH A PLOT OF MTTF VS. FAILURES? ',/)
      END
```

SUBROUTINE RANNOS

```
      real function rannos(seed)
c          The function RANNOS is a pseudo random number generator
c     based on congruential methods involving modulo arithmetic.
c     The multiplication congruential procedure
c
c          R    = K R  (modulo M)
c           n+1      n
c
c     is the recursive algorithm coded below.  K = 65539 is the
c     prime number and the modulo is m = 2 ** 31.
c
c          The initial input to RANNOS, the seed, is a positive
c     odd integer value of up to 9 digits in length.  (If these
c     conditions are not met, then RANNOS will force the value
c     to be positive and odd).
c
c          Computataions within RANNOS are d.p. floating point
c     to maintain precision and avoid repitition within the generated
c     random number cycle.
c
      integer seed,old
      real*8  rold,two31,two311,a,b,c,rem
      logical*1 normal /.false./
c
c     on the first entry into RANNOS, the seed is used to compute
c     the first random number.
c
      if (normal) go to 10
      normal = .true.
c
c     make sure the seed is positive and odd
c
      old = (seed/2)*2 + 1
c
c     convert the seed to floating point and create 2**31
c     and 2**31 -1 for this computer.
c
      rold = old
      two31 = 2. ** 31
      two311 = two31 - 1.
c
c     compute R(N) * K
c
 10   rold = rold * 65539
c
c     If the computed value is less than 2**31, remaindering
c     is not necessary.
c
      if (rold .lt. two31) go to 18
c
c     Remaindering is necessary.  The division of A by2,
c     the multiplication  of I by 2, and the retention of
c     odd or even are done in this way to prevent machine
c     arithmetic overflow on 32 bit word machines.
c
      a = rold / two31
      i = a / 2
      c = i * 2
c
c     was the original value odd or even?
c
      if ((c+1) .gt. a) go to 14
c
c     original was odd
c
      a = c + 1
      go to 16
c
c     original value was even
c
 14   a = c
 16   b = a * two31
      rem = rold - b
      go to 20
 18   rem = rold
 20   continue
c
c     the result of the division of the remainder by 2**31-1 is
c     the new random number.  The remainder is saved for the
c     seed when RANNOS is again called durring the present run.
c
      rannos = rem / two311
      rold = rem
      return
      end
```

SUBROUTINE HISTOGRAM

```
      SUBROUTINE HSTGRM(ARRAY,DELTA,TYPE)
C
C     HSTGRM PLOTS A HISTOGRAM OF THE DATA PASSED FROM THE STAT
C     PROGRAMS.  THE PLOT IS 110 CHARACTERS / LINE AND 50 LINES.
C     THE CALLING ROUTINE PASSES THREE ARGUMENTS:
C
C         ARRAY - CONTAINS THE USER HISTOGRAM DATA
C         DELTA - CONTAINS THE BUCKET SIZE, I.E. THE PERCENTAGE
C                 FACTOR THAT USED IN FILLING THE LOCATIONS OF
C                 USER ARRAY
C         TYPE - USER DESIRED TITLE OF THE HISTOGRAM
C
C     HSTGRM TRANSFERS THE USER INPUT TO LOCAL STORAGE BECAUSE
C     ITS INTERNAL ARRAY 'HISTO' IS DESTROYED IN THE PRINTING
C     PROCESS.
C       IN ADDITION, HSTGRM ASSUMES THE FOLLOWING: 'ARRAY HAS A
C     MAXIMUM OF 100 LOCATIONS; THE DATA IS ORDERED IN 'ARRAY';
C     USING DELTA AS AN INDEX; ALL DATA IS POSITIVE.
C
      CHARACTER * 1 C(100)
      CHARACTER * 1 BLANK /' '/
      CHARACTER * 10 VERT
      CHARACTER * 10 TICK /'         -'/
      CHARACTER * 10 SCALE (11)  /'    100  +',
     1 '     90  +','     80  +','     70  +',
     2 '     60  +','     50  +','     40  +','     30  +',
     3 '     20  +','     10  +','      0  +'/
      CHARACTER * 14 TYPE
      CHARACTER *6 ZERO /'  0.00'/
      CHARACTER * 1 HDR (606)
      CHARACTER * 1 LINE1 (101), LINE2 (101), LINE3 (101),
     1              LINE4 (101), LINE5 (101), LINE6 (101)
      REAL HISTO(100),DELTA
      REAL LARGE,SUM
      REAL PERCENT,ARRAY(100)
      INTEGER B,BL
      REAL RHDR(100)
C
C     ARRAY - USER HISTOGRAM DATA
C     B  - NUMBER OF BUCKETS OR HISTOGRAM COLUMNS
C     BL - SIZE OF A BUCKET (WIDTH OF A HISTOGRAM COLUMN)
C     BLANK - THE SPACE OR BLANK CHARACTER
C     C   - 100 CHARACTER PRINT AREA FOR HISTOGRAM COLUMNS
C     DELTA - HISTOGRAM CLASS INTERVAL OR BUCKET SIZE (PERCENTAGE0
C     HISTO - INTERNAL HISTOGRAM ARRAY
C     LARGE - LARGEST VALUE IN THE HISTO ARRAY
C     LINE1--LINE6  ARE STORAGE AREAS IN WHICH THE HORIZONTAL HEADER
C                   OF THE HISTOGRAM ARE CREATED PRIOR TO PRINTING
C     PERCENT - COMPUTED VALUE BETWEEN 0 AND 100%, USED TO
C               DETERMINE COLUMNAR HEIGHT OF A HISTO VALUE
C     SCALE - VERTICAL PERCENTAGE SCALE HEADER DATA OF THE HISTOGRAM
C     SUM - SUM OF ALL VALUES IN HISTO
C     TICK - A DASH USED ON THE VERTICAL SCALE PRINTING
C     TYPE - USER DESIRED HISTOGRAM TITLE
C     VERT - THE FIRST 10 LOCATIONS OF AN OUTPUT PRINT LINE, HOLDS
C            THE % AND TICK MARKS FOR VERTICAL SCALING
C
C
C     TRANSFER THE USER ARRAY TO HISTO LOCAL STORAGE
C
      DO 5 I = 1,100
      HISTO (I) = ARRAY(I)
 5    CONTINUE
C
C             CLEAR THE PRINT LINES TO BLANKS
C
      DO 10 I = 1,100
      C(I) = BLANK
      LINE1(I) = BLANK
      LINE2(I) = BLANK
      LINE3(I) = BLANK
      LINE4(I) = BLANK
      LINE5(I) = BLANK
      LINE6(I) = BLANK
 10   CONTINUE
C
C     CREATE THE REAL NOS FOR THE HISTOGRAM HEADER USING DELTA
C
      DO 12 I = 1,100
      RHDR (I) = DELTA * I
 12   CONTINUE
```

```
C
C       TO EFFICIENTLY CONVERT RAL NOS TO THEIR CHARACTER REPRESENTATION
C       THE VALUES FOR THE HEADER DATA (REAL NOS) ARE WRITTEN TO A
C       TEMPORARY FILE UNDER FORMAT CONTROL, AND THEN READ BACK IN TO A
C       CHARACTER ARRAY FROM WHIH HEADER PRINT LINES ARE CREATED.
C
        OPEN (UNIT=08,RECL=608,NAME='CHAR.DAT',STATUS='UNKNOWN')
        WRITE (08,1010)ZERO,(RHDR(I),I=1,100)
        ENDFILE 08
        CLOSE (08,STATUS='SAVE')
        OPEN (UNIT=08,NAME='CHAR.DAT',STATUS='UNKNOWN')
        REWIND 08
        READ (08,1050)(HDR(J),J=1,606)
        CLOSE (08,STATUS='DELETE')
C
C       CREATE THE HORIZONTAL HEADER LINES FOR THE HISTOGRAM.
C
        N = 1
        DO 15 I = 1,101,5
        LINE1 (I) = HDR (N)
        LINE2 (I) = HDR (N+1)
        LINE3 (I) = HDR (N+2)
        LINE4 (I) = HDR (N+3)
        LINE5 (I) = HDR (N+4)
        LINE6 (I) = HDR (N+5)
        N = N + 30
 15     CONTINUE
C
C       PRINT THE HEADER AND THE TICK MARKS
C
        WRITE (06,1000)TYPE
        WRITE (06,4000)LINE1
        WRITE (06,4000)LINE2
        WRITE (06,4000)LINE3
        WRITE (06,4000)LINE4
        WRITE (06,4000)LINE5
        WRITE (06,4000)LINE6
        WRITE (06,2500)
C
C       DELTA WILL DETERMINE THE NUMBER OF BUCKETS AND THEIR LOCATION
C       WITHIN A PRINT LINE.  IE, IF DELTA = .01, THEN THERE ARE 100
C       BUCKETS (100%) AND EACH BUCKET IS THE SAME AS ONE PRINT LOCATION
C       ON THE LINE.  IF DELTA = .05, THERE WILL BE 100 BUCKETS, EACH
C       BUCKET REPRESENTING AN INTERVAL OR DELTA OF .05
C
        B = 1. / DELTA
        BL = DELTA * 100
C
C       SCAN THE INPUT ARRAY (ALL 100 LOCATIONS) TO FIND THE LARGEST
C       VALUE (LARGE) AND THE SUM OF ALL VALUES (SUM) IN THE ARRAY.
C
        LARGE = 0
        SUM = 0
        DO 20 II = 1,100
        SUM = SUM + HISTO (II)
        IF (HISTO(II) .LT. LARGE) GO TO 20
        LARGE = HISTO (II)
 20     CONTINUE
C
C       ESTABLISH THE PRINTING PARAMATERS WHICH WILL BE USED IN
C       OUTPUTTING THE HISTOGRAM
C       SET UP TO PRINT THE 50 LINES OF THE HISTOGRAM
C
        K = 0
        DO 500 I = 1,50
        PERCENT = SUM * ((100.0001 - (2.*I)) / 100.0)
C
C       IS THE VERTICAL SCALE NEEDED FOR THIS PRINT LINE, IE
C       K = 0,5,10,15,20,-----50
C
        IF (K .NE. 0) GO TO 50
C
C       TAKE CARE OF THE FIRST PRINT LINE FOR SCALE AS A SPECIAL
C       CASE.
C
        M = 1
        K = K + 1
        GO TO 60
```

```
C
C
 50     IF (MOD(K,5) .NE. 0) GO TO 100
C
C       YES, INSERT PERCENTILE AND + IN COLUMNS 1-10
C
        M = M + 1
 60     VERT = SCALE (M)
        K = K + 1
        GO TO 110
C
C       NO SCALE MARKING IS NEEDED FOR THIS PRINT LINE
C
 100    VERT = TICK
        K = K + 1
C
 110    CONTINUE
C
C       FOR EACH PRINT LINE, SCAN HISTO TO DETERMINE IF A BUCKET
C       WILL CAUSE PRINT LOCATIONS TO BE SET TO NON-BLANK.
C
        DO 200 JK = 1,100
C
C       IF THE BUCKET IS LESS THAN THE MAGNITUDE FOR THIS PRINT LINE
C       IGNORE IT.
C
        IF (HISTO(JK) .LT. PERCENT) GO TO 200
C
C       THE VALUE IS SUFFICIENTLY LARGE TO MODIFY THE PRINT LINE
C
C
C       BASED ON JK, DETERMINE THE LOCATION OF THE BUCKET ON THE PRINT
C       LINE AND SET ITS LOCATION TO AN ASTERIK.
C
        C(JK) = '*'
C
C       CLEAR HISTO(JK) TO ZERO TO AVOID EXTRA PROCESSING
C
        HISTO (JK) = 0.
 200    CONTINUE
C
C       WRITE THE DATA LINE
C
        WRITE (06,2000)VERT,C
C
C
C
 500    CONTINUE
C
C       WRITE THE BOTTOM LINE OF TICK MARKS
C       AND THEIR SUB HEADINGS
C
        WRITE (06,3000)
        WRITE (06,4000)LINE1
        WRITE (06,4000) LINE2
        WRITE (06,4000) LINE3
        WRITE (06,4000) LINE4
        WRITE (06,4000) LINE5
        WRITE (06,4000) LINE6
C
C       ********** ****** FORMAT STATEMENTS **********************
C
 1010   FORMAT (A6,100(F6.2))
 1020   FORMAT (50(F6.2))
 1050   FORMAT (606A1)
 1060   FORMAT (300A1)
 1000   FORMAT (1X,////,8X,20X,A14,' HISTOGRAM'//)
 2000   FORMAT (1X,A10,100A1)
 2500   FORMAT (1X,'    TIME    ',20('+----'))
 3000   FORMAT (1X,10X,20('+----'))
 4000   FORMAT (1X,10X,101A1)
        RETURN
        END
```

SUBROUTINE TPLOT

```
      SUBROUTINE TPLOT (F,M,THETA)
C     TPLOT PLOTS A X/Y GRAPH OFTHE FAILURE DATA PASSED FROM STAT10.
C     THE PLOT IS PRODUCED ON A SINGLE PRINTER PAGE (50 LINES BY
C     100 CHARACTERS PER LINE).
C     THE CALLING ROUTINE PASSES THE FOLLOWING ARGUMENTS:
C
C       F     - ARRAY CONTAINING THE # OF FAILURES PER SAMPLE
C       M     - ARRAY CONTAINING THE MTTF OF EACH SAMPLE
C       THETA - THE COMPUTED THETA VALUE
C
C     TPLOT TRANSFERS THE F AND M ARRAYS TO INTERNAL STORAGE BECAUSE
C     ITS INTERNAL MTTF ARRAY IS DESTROYED DURING PLOT CREATION.  F
C     M, AND THETA ARE ASSUMED TO CONTAIN POSITIVE VALUES GREATER
C     THAN OR EQUAL TO ZERO.
      REAL FAIL(100),MTTF(100),THETA,F(100),M(100)
      REAL MAXF, MAXM
      REAL YSCALE, Y(21)
      REAL MAG
      INTEGER LOC
      INTEGER TFLAG /0/
      INTEGER NENTRY
      INTEGER XSCALE,X(21)
      CHARACTER * 1 CHR
      CHARACTER * 1 XLINE(84)
      CHARACTER * 10 YLINE (11)
      CHARACTER * 1 L1(101), L2(101), L3(101), L4(101)
      CHARACTER * 10 VERT (11)
      CHARACTER * 10 TICK /'         -'/
      CHARACTER * 10 YHEAD
      CHARACTER * 1 C(100)
      CHARACTER *1 PLUS /'+'/
      CHARACTER * 1 BLANK /' '/
C
C     BLANK - THE SPACE OR BLANK CHARACTER
C     C  -  100 CHARACTER PRINT AREA FOR PLOT PRINT LINES
C     CHR - TEMP LOCATION USED TO HOLD AN * OR M WHICH IS USED TO
C           REPRESENT A PLOT POINT ('*' INDICATES A SINGLE VALUE AT
C           THE PRINT LOCATION, 'M' INDICATES MULTIPLE POINTS
C           PLOTTED TO THE SAME LOCATION.
C     F  -  USER PASSED ARRAY CONTANING # OF FAILURES
C     FAIL - TPLOT INTERNAL ARRAY FOR 'F'
C     LOC - HORIZONTAL LOCATION OF PLOT POINT ON THE PRINT LINE C
C     L1 - L4  - STORAGE AREAS IN WHICH HORIZONTAL HEADER DATA IS
C                IS CREATED FOR PRINTING
C     M  -  STAT10 INPUT OF MTTF DATA
C     MAG - MAGNITUDE OF THE Y AXIS (= YSCALE * 50)
C     MAXF - LARGEST VALUE IN ARRAY FAIL
C     MAXM - LARGEST VALUE IN ARRAY MTTF
C     MTTF - INTERNAL MTTF ARRAY
C     NENTRY - # OF ENTRIES IN FAIL AND MTTF
C     PLUS - THE CHARACTER '+'
C     TFLAG - FLAG USED TO DETERMINE WHEN TO PRINT THE THETA LINE
C     THETA - STAT10 THETA VALUE
C     TICK - VERTICAL SCALE MARK ON THE OUTPUT LINE
C     VERT - 10 CHAR PRINT LOCATION USED PRINTING Y AXIS HEADERS
C     X  -  INTEGER ARRAY OF X AXIS HEADER VALUES
C     XLINE - CHARACTER ARRAY OF X AXIS HEADER VALUES
C     XSCALE - SCALING FACTOR FOR EACH PLOT INCREMENT ON THE X AXIS
C     Y  -  REAL ARRAY OF Y AXIS HEADER VALUES
C     YLINE - CHARACTER ARRAY OF Y AXIS HEADER VALUES
C     YSCALE - SCALING FACTOR FOR EACH PLOT INCREMENT ON THE Y AXIS
C
C     INIT VARIABLES
C
      NENTRY = 0
      MAXF = 0.
      MAXM = 0.
```

```
C
C       CLEAR L1 THRU L4 TO BLANKS
C
        DO 20 I = 1,101
        L1(I) = BLANK
        L2(I) = BLANK
        L3(I) = BLANK
        L4(I) = BLANK
 20     CONTINUE
C
C       TRANSFER THE # OF NON-ZERO FAILURE ITEMS AND THEIR ASSOCIATED
C       MTTF VALUES TO INTERNAL ARRAYS.  DURING THE TRANSFER, DETERMINE
C       THE LARGEST VALUE OF FAILURES AND MTTF AS WELL AS THE # OF
C       ENTRIES.
C
        J = 0
        DO 100 I = 1,100
        IF (F(I) .LT. 1) GO TO 100
C
C       PRESET "C" TO BLANKS
C
        C(I) = BLANK
        J = J + 1
        FAIL (J) = F(I)
        MTTF (J) = M (I)
        NENTRY = NENTRY + 1
        IF (FAIL(J) .GT. MAXF) MAXF = FAIL (J)
        IF (MTTF(J) .GT. MAXM) MAXM = MTTF(J)
 100    CONTINUE
C
C       THE SIZE FOR A PLOT OF THE # OF FAILURES VERSUS MTTF IS
C       ONE PRINTER PAGE (50LINES BY 100 CHARACTERS PER LINES).
C       USING MAXF AND MAXM, DETERMINE TEH SCALING FACTORS AND
C       THEN CREATE THE HORIZONTAL AND VERTICAL HEADERS FOR THE
C       PLOT.  THE # OF FAILURES IS PLOTTED ON THE X OR HORIZONTAL
C       AXIS AND THE MTTF IS THE Y OR VERTICAL AXIS.
C
C       CREATE THE HEADER LINES FOR THE HORIZONTAL AXIS.  BECAUSE OF
C       THE POTENTIAL SPREAD  IN THE # OF FAILURES (FROM 1 TO N, WHERE
C       N CAN BE GREATER THAN 1000) TPLOT ASSUMES A MAXIMUM OF
C       9999 FAILURES.  SINCE THERE ARE 100 PHYSICAL LOCATIONS ON THE
C       PRINT LINE, 100 IS USED AS THE ROUNDING FACTOR TO DETERMINE
C       PLOTTING SCALE.
C
        XSCALE = (MAXF + 100) / 100
C
C       SIMILIARLY CREATE THE HEADERS FOR THE Y AXIS.  THE MAX VALUE
C       OF MTTF IS ASSUMED TO BE 999.9999 AND SINCE THERE ARE FIFTY
C       PRINT LINES, 50 IS USED AS THE ROUNDING FACTOR.
C
        YSCALE = (MAXM * 1.1) / 50
C
C
C       IN ORDER TO CONVERT REAL DATA TO CHARACTER FORMAT, AN ARRAY
C       OF THE HEADER DATA IS CREATED USING THE XSCALE AND YSCALE
C       ARRAYS; THE ARRAYS ARE WRITTEN TO A SCRATCHFILE UNDER FORMAT
C       CONTROL; AND THE DATA READ BACK FROM THE FILE INTO A CHARACTER
C       ARRAY.
C
        X(1) = 0.
        DO 150 I = 1,20
        X(I+1) = XSCALE * I * 5
 150    CONTINUE
C
```

```
      Y(1) = 0.
      DO 175 I = 1,10
      Y(I+1) = YSCALE * I * 5
 175  CONTINUE
C
      OPEN (UNIT=08,RECL=130,NAME='CHAR.DAT',STATUS='UNKNOWN')
      WRITE (08,1000) (X(K),K=1,21)
      WRITE (08,1025) ((BLANK,Y(K),PLUS), K=1,11)
      ENDFILE 08
      CLOSE (08,STATUS='SAVE')
      OPEN (UNIT=08,NAME='CHAR.DAT',STATUS='UNKNOWN')
      REWIND 08
      READ (08,1050) XLINE
      READ (08,1075)YLINE
      CLOSE (08,STATUS='DELETE')
C
C     CREATE THE PRINT LINES FOR THE HORIZONTAL AXIS
C
      N = 1
      I = 1
      DO 200 K = 1,21
      L1(I) = XLINE (N)
      L2(I) = XLINE (N+1)
      L3(I) = XLINE (N+2)
      L4(I) = XLINE (N+3)
      N = N+4
      I = I + 5
 200  CONTINUE
C
C     CREATE THE HEADER PORTION FOR THE VERTICAL AXIS
C
      DO 250 I = 1,11
      VERT (12-I) = YLINE (I)
 250  CONTINUE
C
C     WRITE THE TITLE
C
      WRITE (06,2000)
      WRITE (06,2010)
C
C     ESTABLISH A MAGNITUDE FOR THE Y SCALE PRINTING
C
      MAG = MAXM  * 1.1
C
C     PROCESS THE FAILURE DATA AND PRINT THE PLOT
C
      K = 0
      DO 800 I = 1,50
C
C     IS THE VERTICAL SCALE NEEDED ON THEIS PRINT LINE, IE
C     K=0,5,10,15,20,-----50
C
      IF (K .NE. 0) GO TO 350
C
      MM  = 1
      K = K + 1
      GO TO 360
C
 350  IF (MOD(K,5) .NE. 0) GO TO 400
C
C     YES INSERT SCALE VALUE AND + IN COLUMNS 1-10
C
      MM = MM + 1
 360  YHEAD = VERT(MM)
      K = K + 1
      GO TO 410
C
C     SCALE NOT REQUIRED, INSERT TICK MARKS
C
 400  YHEAD = TICK
      K = K + 1
 410  CONTINUE
C
C     DETERMINE THE MAGNITUDE OF THE Y AXIS FOR THIS PRINT LINE
C
      MAG = MAG - YSCALE
C
C     FOR EACH PRINT LINE DETERMINE IF THE MAGNITUDE OF A VALUE IN
C     IN THE MTTF ARRAY IS GREATER THAN THE SCALE OF THE PRINT LINE.
C     IF SO, SET THE PRINT LOCATION ASSOCIATED WITH THE VALUE IN
C     FAIL TO AN ASTERICK.
C
      DO 600 JK = 1,NENTRY
C
C     IGNORE THE MTTF VALUE IF IT IS LESS THAN THE PRINT LINE MAG
C
```

```
      IF (MTTF(JK) .LT. MAG) GO TO 450
C
C     THE VALUE IS SUFFICIENTLY LARGE TO MODIFY THE PRINT LINE.
C     BASED ON THE ASSOCIATED VALUE IN FAIL, DETERMINE THE LOCATION
C     ON THE PRINT LINE TO SET TO AN ASTERICK.
C
      CHR = '*'
      LOC = FAIL(JK) / XSCALE + 1
C
C     IF THIS LOCATION IS ALREADY NON-BLANK, SET CHR TO AN 'M'
C     MEANING MULTIPLE POINTS AT THIS LOCATION.
C
      IF (C(LOC) .NE. BLANK) CHR = 'M'
      C(LOC) = CHR
C
C     CLEAR THE MTTF VALUE TO 0 TO AVOID FUTURE PROCESSING
C
      MTTF (JK) = 0.
C
C     BASED ON THE THETA VALUE INPUT FROM THE CALLING ROUTINE
C     A HORIZONTAL LINE WILL BE PRINTED IN THE MIDDLE OF THE PLOT.
C     IF THETA IS ALREADY PRINTED SKIP THIS CODE
C
 450  IF (TFLAG .NE. 0) GO TO 600
      IF (THETA .LT. MAG) GO TO 600
      TFLAG = 1.
C
C     SET UP THE THETA PRINT LINE
C
      YHEAD = '  THETA  -'
C
C     REPLACE ALL BLANK CHARACTERS ON THE LINE WITH A BLANK
C
      DO 500 IJK = 1,100
      IF (C(IJK) .EQ. BLANK) C(IJK) = '-'
 500  CONTINUE
C
 600  CONTINUE
C
C     WRITE THE PRINT LINE
C
      WRITE (06,3000) YHEAD,C
C
C     CLEAR THE DATA LINE FOR THE NEXT PASS
C
      DO 650 KL = 1,100
      C(KL) = BLANK
 650  CONTINUE
C
 800  CONTINUE
C
      WRITE (06,3100)
      WRITE (06,3200)L1
      WRITE (06,3200)L2
      WRITE (06,3200)L3
      WRITE (06,3200)L4
      WRITE (06,4000)XSCALE,YSCALE,THETA
C
C     ***************************  FORMAT STATEMENTS ******************
C
 1000 FORMAT (2114)
 1025 FORMAT (11(A1,F8.4,A1))
 1050 FORMAT  (84A1)
 1075 FORMAT (11A10)
 2000 FORMAT (20X,'PLOT OF BEHAVIOUR OF MTTF EXTIMATE VERSUS THE'
     1 ' NUMBER OF FAILURES',//)
 2010 FORMAT (' MTTF EST')
 3000 FORMAT (1X,A10,100A1)
 3100 FORMAT (1X,10X,20('+----'),'+')
 3200 FORMAT (1X,10X,101A1)
 4000 FORMAT (1X,//,' HORIZONTAL SCALE ',I4,/,
     1  ' VERTICAL SCALE ',F8.4,/,
     2  ' THETA ',F8.4//)
      RETURN
      END
$
```

# CHAPTER 11

## Covariance Analysis

CHAPTER 11

# Covariance Analysis

## Summary

Previous chapters (2 and 9) have discussed analysis of variance and regression analysis. Covariance analysis is a mixture of the two, where a model includes both continuous and discrete variables. The method presented here may also be adapted to give a numerical solution to analysis-of-variance type problems (as was promised in concluding Chapter 2). An example is given for the ESD data used earlier in Chapter 2.

## Introduction

The usual regression analysis fits some dependent variable Y on observed values of independent variables $X_1$, $X_2$, ...$X_n$. In reliability work, Y is often a failure rate of the $X_i$ might be, for example, temperature (measured in $^oC$) and hours of operation. But what if the $X_i$ are not measurable on a continuous scale but are instead discrete (or qualitative)? Application environment is such an example: we cannot relate "Ground, Benign" to "Airborne, Inhabited" numerically although one is known to be worse than the other. The data simply fall into one or another such category. However, what can be done is to derive a numerical realization of the two environments with respect to their effect on failure rate.

## Model

The usual linear model for regression is given by:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \epsilon \qquad (44)$$

where

$\beta$s are to be solved using least squares optimization.

For the covariance model, Y is still observed (e.g., failure rate data) but some of the X variables are now discrete. Suppose, for the moment, that there is just one discrete variable, the first one. Using the environment example, $X_1$ may be "Ground, Benign" environment or "Airborne, Inhabited" environment. Thus, $X_1$ may be set equal to 0 for the former and 1 for the latter:

$$X_1 = \begin{cases} 0 & \text{Ground, Benign} \\ 1 & \text{Airborne, Inhabited} \end{cases} \tag{45}$$

Least squares may not be applied in the usual way, and $\hat{\beta}_1$ (the solution for $\beta_1$) is the numerical realization of the effect of the Airborne environment on failure rate.

Thus, from (45) a model for Airborne failure rates would be given by

$$Y = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 X_2 + \ldots \; n X_n + \epsilon \tag{46}$$

and for the Ground environment the model would be

$$Y = \hat{\beta}_0 + \hat{\beta}_2 X_2 + \ldots \hat{\beta}_n X_n + \epsilon \tag{47}$$

It does not matter whether we code this way or by using 1 for Airborne, 0 for Ground environment: the solution will be the same either way.

Notice that the effect of Ground environment is lost in the constant ($\beta_0$). This does not matter here, but in some cases (with more discrete variables) it may. To remove the difficulty, it may be preferable to define $X_1$ as:

$$X_1 = \begin{cases} -1 & \text{Ground, Benign} \\ 1 & \text{Airborne, Inhabited} \end{cases} \quad (48)$$

(Though the solution is essentially the same, it may be more convenient this way.)

But suppose it was necessary to analyze more than two environments. Then it is necessary to introduce more than one X variable to solve. Generally, if there are p cases of a discrete variable, we need (p - 1) X variables to accommodate it. Thus, introducing a third environment (Naval) we now require (3 - 1) = 2 "X" variables defined as:

$$X_1 = \begin{cases} 1 & \text{Airborne, Inhabited} \\ 0 & \text{otherwise} \end{cases} \quad (49)$$

$$X_2 = \begin{cases} 1 & \text{Naval} \\ 0 & \text{otherwise} \end{cases} \quad (50)$$

Thus the following table of dummy variable coding results:

| | $X_1$ | $X_2$ |
|---|---|---|
| Ground, Benign | 0 | 0 |
| Airborne, Inhabited | 1 | 0 |
| Naval | 0 | 1 |

The optional use of -1 instead of 0 applies as before.

The principle may be extended to any number of cases of a given discrete variable (provided there is enough data to analyze them all).

The continuous variables are also included in the usual way according to the value they take.

Transformations

The model used (44) assumes the linear additive model, which may not be reasonable. For example, if the model is multiplicative, a logarithmic transformation is appropriate (as explained in Chapter 9). However, taking the logarithm of $X_1$ in the illustration above could result in log(-1) which is not possible. Thus, coding of discrete variables applies <u>after</u> transformation. The multiplicative model is given (with the variables and coefficients redefined for this model) by:

$$Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} \ldots X_n^{\beta_n} \epsilon \tag{51}$$

Taking logarithms:

$$\log Y = \log \beta_0 + \beta_1 \log X + \beta_2 \log X_2 + \ldots + \beta_n \log X_n + \log \epsilon \tag{52}$$

The discrete variable is now coded in the $\log X_i$. Using the second illustration above, the coding would be:

$$\log X_1 = \begin{cases} 1 & \text{Airborne, Inhabited} \\ 0 & \text{otherwise} \end{cases} \tag{53}$$

$$\log X_2 = \begin{cases} 1 & \text{Naval} \\ 0 & \text{otherwise} \end{cases} \tag{54}$$

Transforming back again, after solution for $\beta_0$, $\beta_1$ and $\beta_2$ would give:

$Y = e^{\beta_0} e^{\beta_1 \epsilon}$ (Airborne, Inhabited)
$Y = e^{\beta_0} e^{\beta_2 \epsilon}$ (Naval)
$Y = e^{\beta_0 \epsilon}$ (Ground, Benign)

## Model Solution

The model, once set up, may be solved using least squares in the same way as straightforward regression-type problems. The same measures are applicable for goodness-of-fit and significance of variables. For example, if the Ground and Airborne environments (used before) do not differ in their respective effects on failure rate, then the $X_1$ variable (which is a dummy variable for environment taking on the values 0 or 1) will be found not significant, as measured by the F-statistic.

## Example

This example is a numerical solution of the ESD data introduced in Chapter 2. (In actual fact a numerical model had not been required for the original study of which the data is a part; however, it was quite a good data set for illustration purposes.)

Recall the data on failure voltages as follows:

TABLE 25: FAILURE VOLTAGE DATA

| | Hex Inverter | | Quad D Flip Flop | |
|---|---|---|---|---|
| | APTTG | APTTF | APTTG | APTTF |
| Tester I | 200 | 1800 | 1800 | 2200 |
| | 200 | 800 | 700 | 1400 |
| | 200 | 700 | 500 | 2200 |
| | 400 | 1800 | 1800 | 2200 |
| | 600 | 1400 | 2200 | 3000 |
| | 500 | 1400 | 2600 | 2200 |
| Tester II | 400 | 3000 | 500 | 4600 |
| | 400 | 2200 | 700 | 4600 |
| | 400 | 2200 | 3400 | 4600 |
| | 300 | 1400 | 300 | 2200 |
| | 400 | 900 | 400 | 3000 |
| | 400 | 900 | 800 | 4600 |

The data may be fitted to a model of the form of (54):

$$V = \beta_0 X_1^{\beta_1} X_2^{\beta_2} X_3^{\beta_3} \epsilon$$

which transforms to:

$$\log V = \log \beta_0 + \beta_1 \log X_1 + \beta_2 \log X_2 + \beta_3 \log X_3 + \log \epsilon$$

where V is failure voltage

$$\log X_1 = \begin{cases} 0 & \text{Tester I} \\ 1 & \text{Tester II} \end{cases}$$

$$\log X_2 = \begin{cases} 0 & \text{Hex Inverter} \\ 1 & \text{Quad D flip flop} \end{cases}$$

$$\log X_3 = \begin{cases} 0 & \text{APTTG} \\ 1 & \text{APTTF} \end{cases}$$

Thus the data of Table 25 may be coded and run though a regression program. The data for input were as in Table 26.

The regression solution found the significant variables ($\alpha$ = 5%) to be $X_2$ and $X_3$ (i.e., component type and test circuit), with F values of 31.3 and 62.8 (far exceeding the critical value at the 5% level of 2.82).

$$\log \beta_0 = 2.575$$
$$\beta_2 = 0.379 \qquad \text{(standard error = 0.0677)}$$
$$\beta_3 = 0.537 \qquad \text{(standard error = 0.0677)}$$

The standard error of $\hat{\beta}_2$ and $\hat{\beta}_3$ was 0.0677. The particular program used did not give standard error for $\hat{\beta}_0$; however, this is also important

TABLE 26: LOGARITHMS OF FAILURE VOLTAGES AND CODED COVARIATES

| $\log_{10}V$ | $\log X_1$ | $\log X_2$ | $\log X_3$ | log V | $\log X_1$ | $\log X_2$ | $\log X_3$ | log V | $\log X_1$ | $\log X_2$ | $\log X_3$ | log v | $\log X_1$ | $\log X_2$ | $\log X_3$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2.301 | 0 | 0 | 0 | 3.255 | 0 | 0 | 1 | 3.255 | 0 | 1 | 0 | 3.342 | 0 | 1 | 1 |
| 2.301 | 0 | 0 | 0 | 2.903 | 0 | 0 | 1 | 2.845 | 0 | 1 | 0 | 3.146 | 0 | 1 | 1 |
| 2.301 | 0 | 0 | 0 | 2.845 | 0 | 0 | 1 | 2.699 | 0 | 1 | 0 | 2.342 | 0 | 1 | 1 |
| 2.602 | 0 | 0 | 0 | 3.255 | 0 | 0 | 1 | 3.255 | 0 | 1 | 0 | 3.342 | 0 | 1 | 1 |
| 2.778 | 0 | 0 | 0 | 3.146 | 0 | 0 | 1 | 3.342 | 0 | 10 | 0 | 3.477 | 0 | 1 | 1 |
| 2.699 | 0 | 0 | 0 | 3.146 | 0 | 0 | 1 | 3.415 | 0 | 1 | 0 | 3.342 | 0 | 1 | 1 |
| 2.602 | 1 | 0 | 0 | 3.477 | 1 | 0 | 1 | 2.699 | 1 | 1 | 0 | 3.663 | 1 | 1 | 1 |
| 2.602 | 1 | 0 | 0 | 3.342 | 1 | 0 | 1 | 2.845 | 1 | 1 | 0 | 3.663 | 1 | 1 | 1 |
| 2.602 | 1 | 0 | 0 | 3.342 | 1 | 0 | 1 | 3.531 | 1 | 1 | 0 | 3.663 | 1 | 1 | 1 |
| 2.477 | 1 | 0 | 0 | 3.146 | 1 | 0 | 1 | 2.477 | 1 | 1 | 0 | 3.342 | 1 | 1 | 1 |
| 2.602 | 1 | 0 | 0 | 2.954 | 1 | 0 | 1 | 2.602 | 1 | 1 | 0 | 3.477 | 1 | 1 | 1 |
| 2.602 | 1 | 0 | 0 | 2.954 | 1 | 0 | 1 | 2.903 | 1 | 1 | 0 | 3.663 | 1 | 1 | 1 |

(The data are arranged as in Table 25 for clarity.)

since some assessment of the likely error in the fitted model is preferable. Note that the interaction between any two variables is ignored by this type of analysis: interactions are assumed zero here. Interactions may be accomodated by coding in extra variables; this is left out for simplicity.

Thus, the solution is given by:

$$\log V = 2.575 + 0.379 \log X_2 + 0.537 \log X_3 + \log \epsilon$$

The error in the fitted model ($\log \epsilon$) is sketched below for all the data, as a percentage.

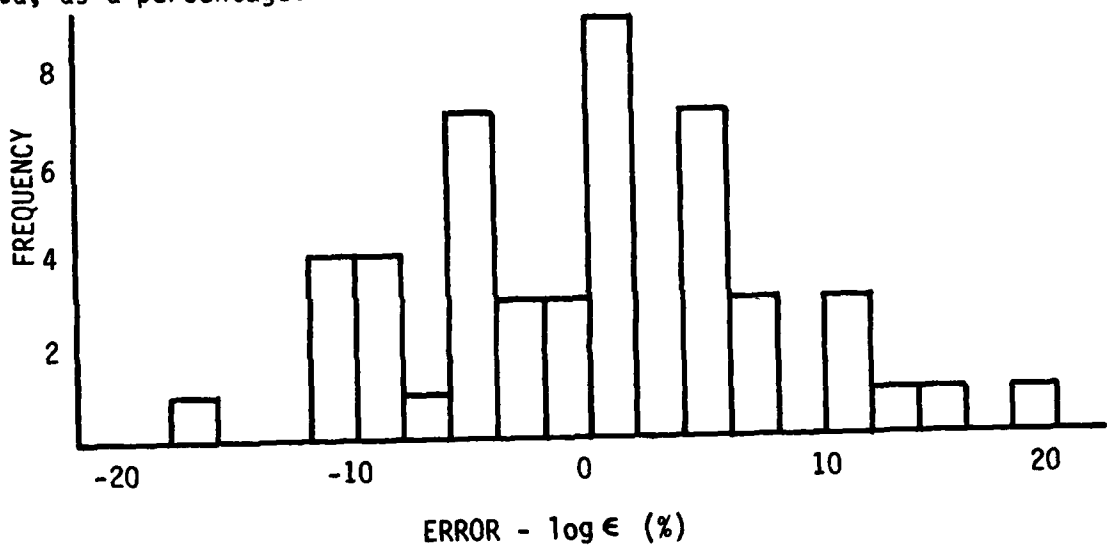



FIGURE 32: RESIDUAL ERROR

We can now readily see that the failure voltage for the two different components differs (nominally) by a factor of $10^{0.379}$, i.e., about 2.39 and also that the two test circuits induce failure voltages which differ by a factor of (nominally) $10^{0.537} \simeq 3.44$.

Thus, the basic model may be written as:

$$V = 375.56 \times \begin{Bmatrix} 1 \text{ for the hex inverter} \\ 2.39 \text{ for the quad D flip flop} \end{Bmatrix} \begin{Bmatrix} 1 \text{ for APTTG} \\ 3.44 \text{ for APTTF} \end{Bmatrix} \times \epsilon$$

$\epsilon$ may be specified also, probably as a lognormal distribution about the fitted model. In practice, the interest is generally for a visual appreciation such as was given in the sketch of residuals though the residuals were plotted on a logarithmic scale ($\log \epsilon$). A final model is shown in Figure 33.

## Review

The general linear model may be used to solve problems in discrete and continuous variables by using dummy variables (taking on values 0 or 1) to represent the discrete variables in usual regression models.

For p cases within a discrete variable, (p - 1) dummy variables are required to fit a model. The model may be used to determine which variables (discrete or continuous) significantly influence the dependent variable. The usual regression statistics apply.

The results found in the example compare well to the results of the analysis of variance Chapter 2 though interactions are ignored here. These results give a numerical solution for the effects of each variable in an analysis-of-variance-type problem. The example used only discrete variables, though the model may be used to solve mixed discrete/continuous variable problems.

Note that the worst fit is found where the interaction (which is ignored here for simplicity) was identified in Chapter 2.

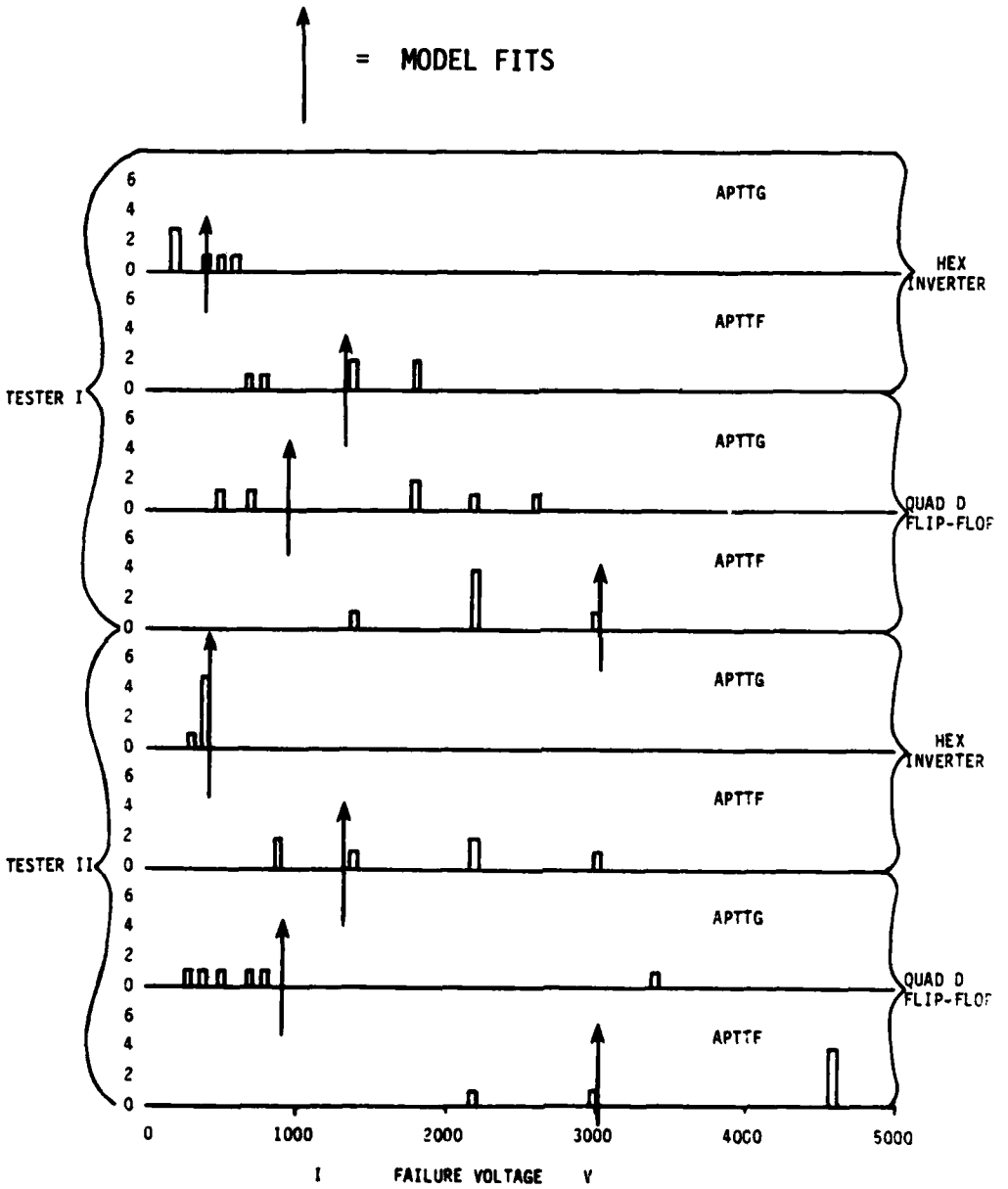



FIGURE 33: MODEL FIT

# CHAPTER 12

## Use of Statistical Tables

CHAPTER 12

# Use of Statistical Tables

## Summary

This chapter briefly discusses the use of statistical tables and points out pitfalls.

## The Need for Tables

Statistical methods make use of numerical evaluation of complex analytical formulae for a variety of probability density functions. To solve these formulae for each application would be time-consuming, and in many cases closed-form solutions are not available. Consequently, over the years a series of standard statistical tables has been compiled with contributions by many workers in the field.

The tables are extensively available; most elementary texts reproduce a set, and there are also dedicated publications (e.g., Ref. 3). This chapter aims to discuss their use and warn against problems which tend to arise in practice.

## What is Tabulated?

Most statistical methods make use of some statistic to objectively evaluate some hypothesis about a set of data. Whether that statistic is F, t, $\chi^2$, Spearman's rank, or some custom-designed statistic, it will have a distribution. The more extreme our value of that statistic the less likely it is that the null hypothesis is true. Thus, statistical tables evaluate given percentiles of a statistical distribution, and we are generally interested in the probability of observing a value as extreme (or more so) as the one observed. "Extreme" can take on either sense (high or low).

Consider a general statistic for the purposes of illustration, which is called $G_S$ and whose distribution is sketched in Figure 34. Statistical tables of $G_S$ would tabulate either the shaded area or the unshaded area. The shaded area gives the probability that $G_S$ will exceed g. If some significance level P is defined in advance then for a one-tailed test we seek g* such that Prob $[G_S > g^*] = P$. g* is then referred to as the critical value. Note that for a two-tailed test we would be interested in both tails, such that Prob $[|G_S| > g^*] = P$ and hence <u>half</u> of P would be allocated to each tail.

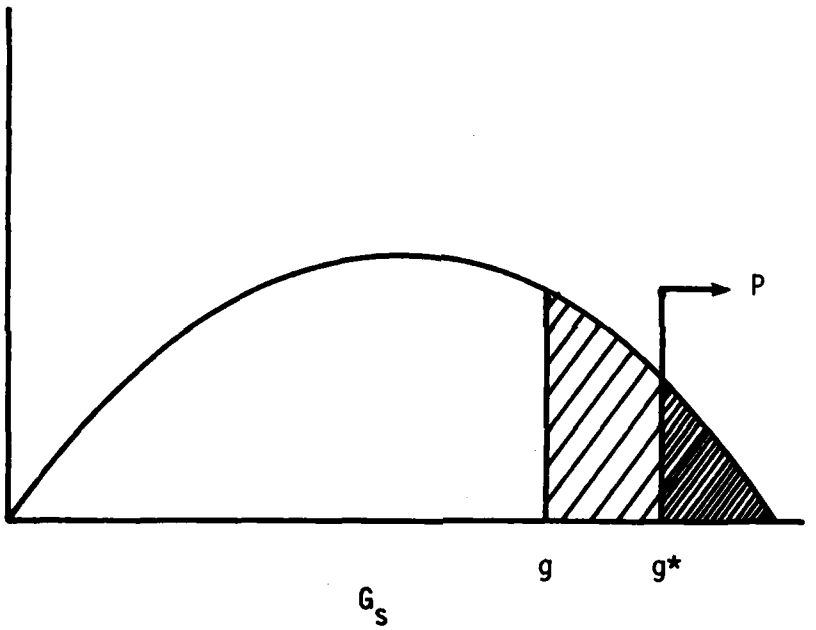



FIGURE 34: A GENERALIZED STATISTIC

Percentiles within an interval may be established by subtraction.

Tables have been constructed from these different viewpoints:

(i) Given $G_S$, the value of P is tabulated
(ii) Given $G_S$, the value of 1-P is tabulated
(iii) Given P, $G_S$ is tabulated

Care is required in establishing which of (i) or (ii) a given table is defined on. For example, chi-square tables have been tabulated either way.

The generalized statistic $G_S$ could be any one of the distributions (F, $\chi^2$, etc.) discussed in this text. Though $G_S$ is shown as continuous, the same principle applies to discrete distributions (e.g., binomial).

In addition, many of the tabulated statistics require the degrees of freedom to be also defined. Such tables include F, $\chi^2$, t, the correlation coefficient, and the Kolmogorov-Smirnov statistic.

Values not given in the tables may be derived crudely by interpolation or by returning to the original formula for the tabulated statistic.

It is interesting to note that many of the major parametric statistics are mathematically related. It has been said (though the author forgets by whom) that if one were to be stranded on a desert island with only a single statistical table, then the choice would be a set of F tables, since from them a number of other tables can be readily derived. The interrelation between some of the major distributions is indicated in Figure 35.

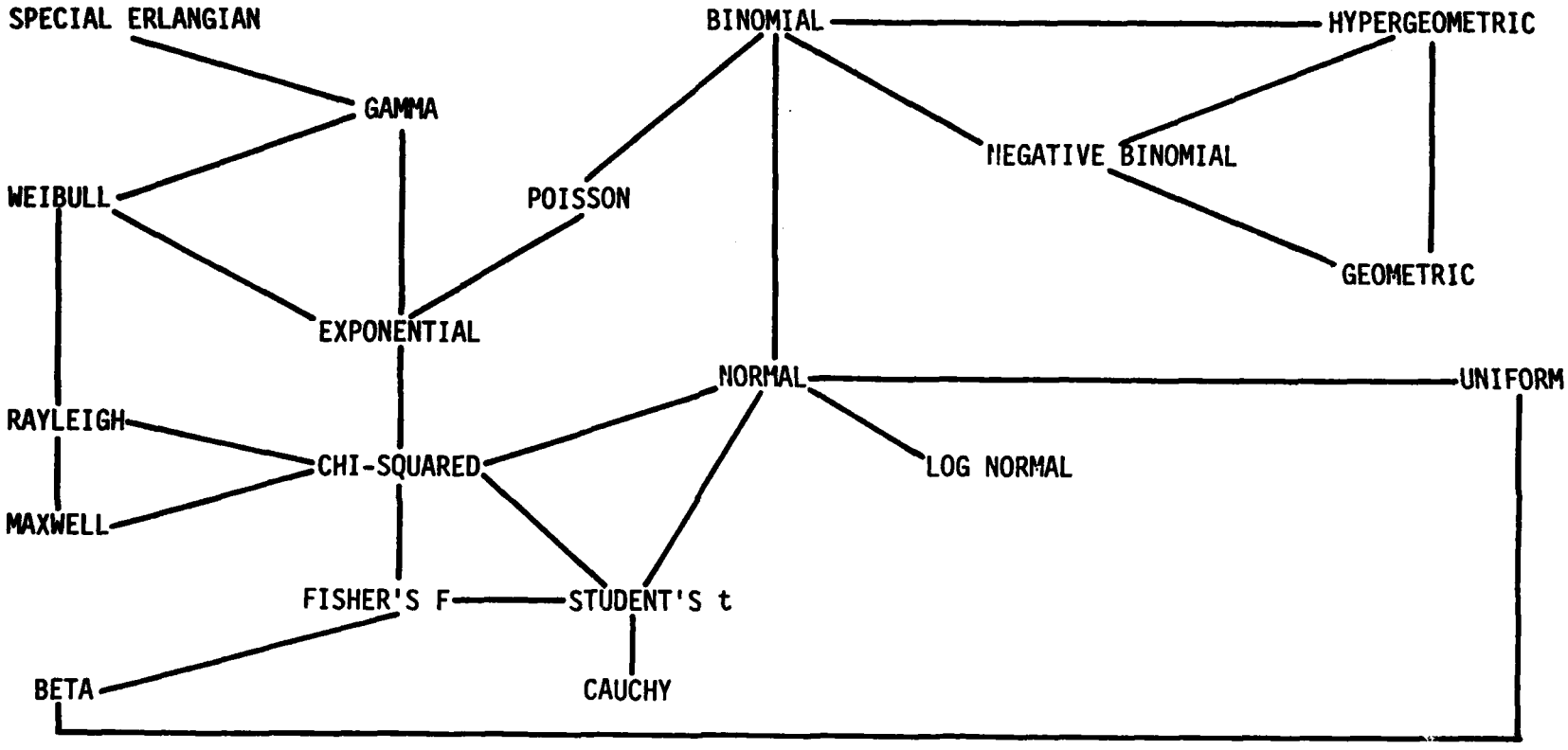



FIGURE 35: THE FAMILY OF DISTRIBUTIONS

# Concluding Remarks

## CONCLUDING REMARKS

This text began by pointing out some limitations of a user guide to applied statistical methods. It is as well to re-emphasize those earlier cautions and conclude with some notes on where to go for the help which may be needed in handling more complex problems or nonstandard data. In many cases, consulting the list of references will provide further depth. The statistical staff at IIT Research Institute are also available for advice or consulting, particularly in the field of reliability data analysis and reliability statistics or models. Supporting data is available from the Reliability Analysis Center (RAC) and Data and Analysis Center for Software (DACS), which are both operated by IIT Research Institute.

For further advice, contact in the first instance


Mr. J.W. Wilbur
Marketing Manager
IIT Research Institute
Reliability Analysis Center
RADC/RAC
Griffiss AFB, NY 13441
Tel: (315) 330-4151


or contact either of the statisticians direct:


Kieron A. Dey
IIT Research Institute
Reliability Analysis Center
Griffiss AFB, Rome, NY 13441
Tel: (315) 330-4151
337-0900

Jorge L. Romeu
IIT Research Institute
Data and Analysis Center for Software
Griffiss AFB, Rome, NY 13441
Tel: (315) 336-0937
336-2359

# References

## REFERENCES


1. Joint Papers of J. Neyman and E.S. Pearson, University of California Press, Berkeley, 1967.

2. Halmos, P., Measure Theory, Van-Nostrand.

3. Murdock, J., and J.A. Barnes, Statistical Tables for Science, Engineering, Management and Business Studies, MacMillan, 1974.

4. Fisher, R.A., Statistical Methods for Research Workers, Hafner, 1954.

5. Hald, A., Statistical Theory with Engineering Applications, Wiley, 1952.

6. Denson, W.K., and K.A. Dey, ESD Susceptibility Testing of Advanced Schottky TTL Devices, Proceedings 1983 EOS/ESD Symposium.

7. Siegel, S., Non Parametric Statistics for the Behavioral Sciences, McGraw-Hill, 1956.

8. Mann, N.R., R.E. Schafer and N.D. Singpurwalla, Methods for Statistical Analysis of Reliability and Life Data. Wiley, 1974.

9. Box, G.E.P., and G.C. Tiao, Bayesian Inference in Statistical Analysis, Addison-Wesley, 1973.

10. MIL-HDBK-217D, Reliability Prediction of Electronic Equipment.

11. Dey, K.A., Correlation of Field Data with Reliability Prediction Models, RADC-TR-81-329.

12. Schafer, R.E., et al. Reliability Growth Study, RADC-TR-75-253.

13. Johnson, L.G., The Statistical Treatment of Fatigue Experiments, Elsevier, 1964.

14. Lehmann, E.H., Nonparametrics - Statistical Methods Based on Ranks, Holden-Dory, 1975.

15. Wetherill, G.B., Sampling Inspection and Quality Control, Barnes and Noble, Inc., 1969.

16. MIL-M-38510, Appendix B, General Specification for Microcircuits.

17. MIL-S-19500, Appendix C, General Specification for Semiconductor Devices.

18. MIL-STD-105D, Sampling Procedures and Tables for Inspection by Attributes.

19. IMSL Library (of Statistical Subroutines), IMSL, Inc.

20. Mage, D.T., An Objective Graphical Method for Testing Normal Distributional Assumptions Using Probability Plots, The American Statistican, Vol. 36, No. 2, May 1982.

21. Bulmer, M.G., Principles of Statistics, Oliver and Boyd, 1967.

22. Keen, J., and D.J. Farlie, The Design and Interpretation of Experiments, GEC-Hirst Research Centre (Internal Report), 1974.

23. Draper, N.R., and H. Smith, Applied Regression Analysis, Wiley, 1966.

24. Box, Hunter and Hunter, Statistics for Experimenters, Wiley, 1978.

25. Sen, P.K., Estimates of the Regression Coefficient Based on Kendall's Tau, JASA, Vol. 63, pp. 1379-1389.

26. Romeu, J.L., and C. Turner, An Investigation of the Effects of Technology on Development Efforts, IIT Research Institute (Internal Report), May 1982.

27. Graybill, F.A., Theory and Application of the Linear Model, Duxbury, 1976.

28. Mosteller, F., and J.W. Tukey, Data Analysis and Regression, Addison-Wesley, 1977.

29. Cox, D.R., Regression Models and Life Tables, J.R.S.S.(A) #2, 1972.

30. Chaterjee and Price, Regression Analysis by Example, Wiley, 1978.

31. Efron, B., The Jacknife, the Bootstrap and Other Resampling Plans, SIAM, 1982.

32. Coppola, A., Bayesian Reliability Tests Made Practical, RADC-TR-81-106, (AD-A108150).

33. Dey, K.A., Statistical Analysis of Noisy, Incomplete Failure Data, 1982 Annual Reliability and Maintainability Symposium Proceedings.

34. Dodge, H.F., and H.G. Romig, Sampling Inspection Tables, Single and Double Sampling, 2nd ed., Wiley, 1959.

# ADDITIONAL RAC SERVICES

**Search Services**

Retrospective Searches are conducted at a flat fee of $125 per search. If no references are identified, a $50 service charge will be made in lieu of the above. For best results, please call or write for assistance in formulating your search question. An extra charge, based on engineering time and costs, will be made for evaluating, extracting or summarizing information from the cited references.

**Consulting Services**

Consulting Service fees are determined by the costs incurred in the conduct of the designed work, including staff time and overhead, materials and other expenses. Work will be initiated upon receipt of a signed purchase order. We will be pleased to prepare firm cost proposals.

**Full Service Participating Plans**

Two plans are offered to both government and industry

| | |
|---|---|
| **Participating Member (PM)**..... | $1,600 |
| **Participating Associate (PA)**..... | 400 |

Services provided to a Participant in either plan are:

- o Automatic receipt of one (1) copy of each RAC microcircuit and semiconductor device databook issued over twelve months at a savings of $70.
- o Availability of additional copies of each of the above databooks at 20% off list price.
- o Discount on registration fees for RAC sponsored training courses, seminars, workshops, etc.

In addition, the **Participating Member** may access RAC resources as needed without issuing purchase orders. Up to 50 man-hours of professional consultation are authorized.

**Blanket Purchase Order**

The Blanket Purchase Order option enables you to write a single Purchase Order for a stipulated maximum dollar amount (depending on your needs) and active time duration (a one-year period is suggested), but you pay only for services rendered or documents purchased.

**Military Agencies:** Blanket Purchase Agreement, DD Form 1155, may be useful for ordering RAC reports and/or services. Please stipulate maximum dollar amount authorized and cutoff date on your order. Also specify services (e.g., publications, search services, etc.) to be provided. Identify vendor as IIT Research Institute (Reliability Analysis Center).

**Ordering Information**

Place orders or obtain additional information directly from the Reliability Analysis Center. Clearly specify the publications and services desired. Except for blanket purchase orders, prepayment is required. All foreign orders must be accompanied by a check drawn on a U.S. bank. Please make checks payable to IITRI/RAC.

## SERVICE FEE SCHEDULE AND ORDERING INFORMATION

**March 1983**

| | | | Issue Date | Price Per Copy Domestic | Foreign |
|---|---|---|---|---|---|
| **Component Reliability Databooks** | | | | | |
| ( ) | MDR-14 | Hybrid Circuit Data | Mar. 1980 | $60.00 | $70.00* |
| ( ) | MDR-15 | Digital Evaluation and Generic Failure Analysis Data - Vols. I and II | Aug. 1980 | 60.00 | 70.00** |
| ( ) | MDR-16 | Linear/Interface Data | Feb. 1981 | 60.00 | 70.00** |
| ( ) | MDR-17 | Digital Failure Rate Data | Aug. 1981 | 60.00 | 70.00** |
| ( ) | MDR-18 | Memory/LSI Data | Feb. 1982 | 60.00 | 70.00** |
| ( ) | DSR-3 | Transistor/Diode Data | Jan. 1980 | 60.00 | 70.00** |
| | | Complete Set: $310 ($360 non-U.S.) | | | |
| ( ) | NPRD-2 | Nonelectronic Parts Reliability Data | Aug. 1981 | 60.00 | 70.00* |
| ( ) | VZAP-1 | Electrostatic Discharge Susceptibility Data | Apr. 1983 | 95.00 | 105.00** |
| **Equipment Databooks** | | | | | |
| ( ) | EERD-1 | Electronic Equipment Reliability Data | Oct. 1980 | 60.00 | 70.00** |
| ( ) | EEMD-1 | Electronic Equipment Maintainability Data | Oct. 1980 | 60.00 | 70.00* |
| **Handbooks** | | | | | |
| ( ) | RDH-376 | Reliability Design Handbook | Mar. 1976 | 36.00 | 46.00** |
| ( ) | MFAT-1 | Microelectronics Failure Analysis Techniques Procedural Guide | July 1981 | 125.00 | 135.00*** |
| **Technical Reliability Studies** | | | | | |
| ( ) | TRS-1 | Microcircuit Screening Effectiveness | | 36.00 | 46.00* |
| ( ) | TRS-2 | Search and Retrieval Index to IRPS Proceedings-1968 to 1978 | | 24.00 | 34.00** |
| ( ) | TRS-3A | EOS/ESD Technology Abstracts | | 36.00 | 46.00* |
| ( ) | SOAR-1 | ESD Protective Materials and Equipment: A Critical Review | | 36.00 | 46.00* |
| ( ) | SOAR-2 | Practical Statistical Analysis for the Reliability Engineer | | 36.00 | 46.00* |
| **Symposium Proceedings** | | | | | |
| ( ) | EOS-1 | Electrical Overstress/Electrostatic Discharge 1979 Symposium Proceedings | | 24.00 | 34.00* |
| ( ) | EOS-2 | Electrical Overstress/Electrostatic Discharge 1980 Symposium Proceedings | | 24.00 | 34.00* |
| ( ) | EOS-3 | Electrical Overstress/Electrostatic Discharge 1981 Symposium Proceedings | | 24.00 | 34.00* |
| ( ) | EOS-4 | Electrical Overstress/Electrostatic Discharge 1982 Symposium Proceedings | | 24.00 | 34.00* |

*For air mail shipment to points outside North and Central America, add $10.00 per item
**For air mail shipment to points outside North and Central America, add $15.00 per item
***For air mail shipment to points outside North and Central America, add $35.00 per item

**Quantity Purchase Discounts** - Discounts (on multiple copies of a single title ordered at one time) are:

| Quantity | Discount | Quantity | Discount |
|---|---|---|---|
| 1-2 | list | 10-19 | 33-1/3% off list |
| 3-6 | 15% off list | 20-49 | 45% off list |
| 6-9 | 20% off list | 50-99 | 60% off list |
| | | 100 or more | negotiable |

---

**ORDER FORM**

Enclosed find $_____

Send order and check to:

**Reliability Analysis Center**
**RADC/RAC**
**Griffiss AFB, NY 13441**

Phone: 315/330-4151 Autovon: 587-4151

Please send me the documents checked above.

Name/Title_____

Organization_____

Address_____

City/State_____Zip_____

**Prepayment of orders is required.** Please make checks payable to IITRI/RAC. Foreign orders must be accompanied by check drawn on a U.S. bank.

***The Reliability Analysis Center is a DoD Information Analysis Center operated by IIT Research Institute, Chicago, IL***